THE USE OF TIME-SERIES METHODS FOR DIFFUSION MODELLING: AN EVALUATION

First-Year Report Submitted To

UNIVERSITY OF CAMBRIDGE

in partial fulfilment of the requirements
of the award of the degree

DOCTOR OF PHILOSOPHY

by

CHRISTOPH E. WEISS

ECONOMICS & POLICY GROUP, JUDGE BUSINESS SCHOOL
UNIVERSITY OF CAMBRIDGE
June 2014

The Use of Time-Series Methods for Diffusion Modelling: An Evaluation

Christoph E. Weiss

June 2014

Abstract

The ability to describe, explain, and predict the diffusion of innovations in a social system is crucial – understanding the dynamic drivers of the diffusion process is a necessity for successful innovation management. This study sets out to evaluate the extant modelling techniques in the field and introduces state-space modelling as a powerful holistic approach to diffusion modelling. A formal theoretical framework for state-space modelling in a diffusion context is provided. The empirical part of the study suggests superiority of a state-space approach as regards description and forecasting of diffusion processes (when compared to the popular Bass and Logistic growth models, as well as ARIMA models) and can also be used to explain such processes well by accommodating regressors and intervention variables in the model framework. Furthermore, we introduce a formal systematic test (within the state-space framework) for the saddle effect that is a feature of many diffusion processes.

Keywords: Innovation Diffusion; Structural Time Series; State-Space Models; Bass Model; Saddle Effect.

1 Introduction

The diffusion of innovations is a key theme in research across many disciplines, inter-alia, economics, marketing, management, communication, sociology, and geography. Innovations span technologies, products, ideas and practices, and it is not surprising that the processes by which innovations spread through populations is of great academic and practical interest. A variety of theoretical models have been developed to explain the working of a number of diffusion mechanisms. Adoption trajectories that arise from different forces, such as for example, contagion, social influence, and social learning, differ in characteristic ways.

In the real world, diffusion trajectories tend to be noisy, and rarely accord neatly with predictions of theoretical models. It is often difficult to identify and measure their features, for example, the "take-off", the reaching of general acceptance, and the existence of saddles along the adoption path. Over and above tractable mechanisms emphasised by models, real trajectories are influenced by mixtures of other time-varying causal forces: in the case of product innovation, by general economic conditions, by innovator strategies (e.g. price and advertising), by the appearance of successor generations of innovations, and also by the heterogeneity (in the population of potential adopters)[1], network structures, as well as other sources of returns to adoption in the population.

Methods of empirical analysis of diffusion paths must characterise observed real world diffusion trajectories precisely. The ability to extract signals from and thereby understand innovation adoption paths is of enormous value from business, policy and academic points of view. In the context of technology, being able to understand and forecast how the adoption of a new technology proceeds should enable policymakers, including national and international regulatory agencies, to improve the design of promotion schemes. In a marketing context, ability to understand and forecast the sales of innovative products should help devising the marketing mix optimally during various stages of the emergence of the innovation. This should help in determining the underlying mechanism and the role of innovation-relevant characteristics.

That is our focus. We establish the great value of state-space models, specifically, structural time series models, in characterising real world diffusion curves. In order to compare different modelling approaches we re-examine the diffusion of tractors in the US from 1910 - 1960. We show that state-space models are able to fit and forecast diffusion curves extremely well, with greater accuracy than the extant models, including the Bass model and the Logistic growth model.[2] We show that a state-space approach is of clear value in identifying saddles in diffusion paths. We establish a systematic inferential approach to test for this feature.

The paper is structured as follows: Section 2 briefly reviews diffusion theory, Section 3 covers diffusion mechanisms and popular extant modelling approaches, Section 4 describes the time-series methods that are seen to be most useful for diffusion modelling, Section 5 describes the data and applies the reviewed extant methods to the diffusion time series, Section 6 introduces a general stepwise procedure that can be used to select state-space models in a diffusion context and applies it to the data, Section 7 evaluates the goodness of fit and the forecasting accuracy of the estimated models, Section 8 introduces explanatory variables and interventions into the state-space models,

[1]The sources of heterogeneity in diffusion processes are manifold and include differences in the costs of adoption, differences in the prior beliefs, and differences in the information-gathering process (Young, 2009).

[2]In this study, we use the basic Bass model as first introduced by Bass (1969). For an overview of the myriad extensions and alterations that have been suggested, see Bass et al. (1994), Shankar and Carpenter (2012). Some of the most important extensions include the Generalized Bass Model (Mahajan et al., 1995), the Virtual Bass Model (Jiang et al., 2006), and the (multi-generational) Norton-Bass Model (Norton and Bass, 1987).

Section 9 reviews former research on the saddle effect and introduces a systematic test for a saddle in the state-space framework, Section 10 concludes.

2 Diffusion Background

A key premise in the innovation diffusion literature is that various forms of heterogeneity among potential adopters can generate variation in their propensities to adopt innovations is (e.g. Rodgers, 2003; Young, 2009; Peres et al, 2010). Rodgers (2003) offers a stylised five-fold classification of the population of potential adopters (in a marketing context, consumers), ranging from the highest propensity to adopt, to the lowest: innovators, early adopters, early majority, late majority, and laggards.[3] The suggestion is that every potential adopter has her unique intrinsic "innovativeness" that determines her time of adoption (Peres et al, 2010) – the later a potential adopter chooses to adopt, evidently, the lower is her revealed interest in the novelty of the product, i.e. her intrinsic innovativeness. In a more general sense, the higher is her risk aversion (Kijek and Kijek, 2010). It should of course be noted that adoption time can also be expected to be powerfully affected by social forces.[4] Hence, the diffusion decision can be conceived as arising from the interplay of innovation and imitation motives.

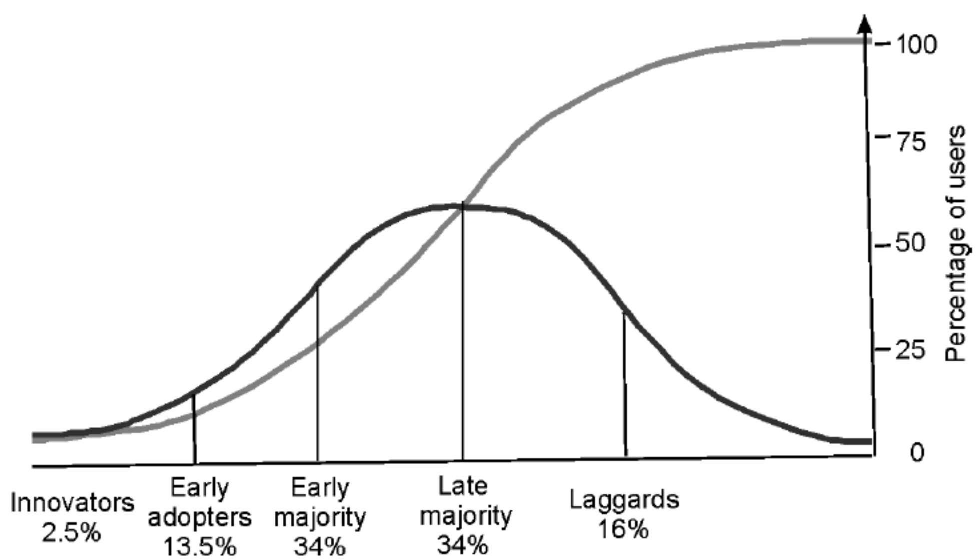

Figure 1: Segmentation of the Population of Potential Adopters (based on Mahajan et al, 2000).

[3]With the proportions: 2.5 %, 13.5 %, 34 %, 34 %, and 16 %, respectively.

[4]For a comprehensive study of this bandwagon effect, see Abrahamson and Rosenkopf (1993, 1997). Other recent elaborate discussions of the impact of social pressure on timing of adoption include Emmanouilides and Davies (2007), Young (2009), Delre et al. (2010), and Cojucaru et al. (2013).

As can be seen in Figure 1, the segmentation of the potential adopters results in a bell-shaped curve for the number of *new adopters*, which translates into an S-shaped curve for the cumulative adoption path. It follows that there are three characteristic phases in the classical diffusion path: After a phase of slow initial uptake, there is a rapid growth phase, which is then followed by the saturation phase.

Another source of heterogeneity that is consistent with the S-shaped diffusion curve would arise from the income distribution (Naseri and Elliott, 2013). Golder and Tellis (1997) show that under the assumption of a log-normal distribution of personal income the diffusion curve is likely to assume an S-shape. Meade and Islam (2006) argue that the affordability of an innovative product rises as its price falls, leading to the S-shaped curve; Van den Bulte and Stremersch (2004) find a positive correlation between the Gini coefficient and the imitation-innovation (q/p) ratio derived from the Bass model.[5]

It is evident that income heterogeneity can play a crucial role in the diffusion of new products. However, the reasoning based on heterogeneity in intrinstic innovativeness and imitation provides a more general basis for our purpose of presenting an empirical modelling approach suitable for characterising the diffusion of *any* innovation. An authoritative definition of the term 'innovation', according to the OECD is: the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations (OECD, 2005). Innovativeness and imitativeness of potential adopters can play a role in the diffusion of all these many types of innovation. Income heterogeneity is primarily applicable to contexts where individual consumers are potential adopters, and is less appealing when the issue is the diffusion of policies, where the potential adopters may be firms, organisations, or countries.

3 Popular Diffusion Modelling Approaches

3.1 Overview of Diffusion Mechanisms

Early research on the diffusion of innovations borrowed models that were used in epidemiology to describe the contagion of infectious diseases. As pointed out by Kijek and Kijek (2010), many commonly used diffusion models can be nested into the following differential equation:

[5] The Bass model and the coefficients mentioned here will be explained in more detail in Section 3.

$$\frac{dN(t)}{dt} = r(t) \times (m - N(t)) \tag{1}$$

where $\frac{dN(t)}{dt}$ is the rate of diffusion; $N(t)$, the cumulative number of adopters at time t; m, the ceiling of potential adopters, and $r(t)$, the coefficient of diffusion.

The coefficient of diffusion, $r(t)$, can be adjusted according to needs and represents the interplay between innovation and imitation forces. A popular systematic classification of the coefficient of diffusion was introduced by Mahajan and Peterson (1985):

- ***External-Influence Model:***

  In the external-influence model, the coefficient of diffusion is a constant, p:

  $$\frac{dN(t)}{dt} = p \times (m - N(t)) \tag{2}$$

  In this setting, the diffusion rate does not depend on the number of prior adopters at time t. It is driven only by intrinsic innovativeness (or risk attitude), and not by imitation, which is why p is also called the coefficient of innovation. Fourt and Woodlock (1960) used the model first, and analyzed innovation diffusion in the grocery market. In general, the external-influence model is a suitable approach when potential adopters are relatively isolated in the social system and where communication is dominantly formalized and hierarchical, i.e. the impact of word-of-mouth is low (Kijek and Kijek, 2010).

- ***Internal-Influence Model:***

  In the internal-influence model, the coefficient of diffusion is $q \times N(t)$:

  $$\frac{dN(t)}{dt} = q \times N(t) \times (m - N(t)) \tag{3}$$

  Here the adoption rate is directly driven by the number of prior adopters, i.e. by imitation, and intrinsic innovativeness does not play a role. The direct interaction between potential adopters and prior adopters (through word-of-mouth, for example) enters the equation through the coefficient of imitation, q. It should be noted that the effect of word-of-mouth is a function of the prior adopters, $N(t)$, and therefore increases with time (Wright et al, 1997; Naseri and Elliott, 2013).[6] The internal-influence model is suitable whenever the group of potential adopters is relatively small and homogenous, or when the legitimation of information prior to adoption is important (Kijek and Kijek, 2010). This form of the internal-influence model takes the form of the Gompertz growth function, which has been used extensively in diffusion research.

- ***Mixed-Influence Model:***

  In the mixed-influence model, the coefficient of diffusion is $p + q \times N(t)$. The model was introduced by Bass (1969), and is a generalization that accommodates both the internal-influence and the external-influence models as special cases. The mixed-influence model is commonly expressed as:

$$\frac{dN(t)}{dt} = (p + \frac{q}{m} \times N(t)) \times (m - N(t)) \tag{4}$$

  This specification allows for both intrinsic innovativeness (through the coefficient of innovation, p) and a word-of-mouth (imitation) effect (through the coefficient of imitation, q). Its flexibility has made the mixed-influence model by far the most frequently used among the three diffusion models that are based on differential equations (Kijek and Kijek, 2010).

It is worth mentioning that in all the discussed models, the ceiling for potential adopters, m, is assumed to be constant. The simplicity of the model is appealing, but as the features of the population through which an innovation diffuses might change over time, this is a limitation. Sharif and Ramanathan (1981) offer a dynamic model in which the ceiling of potential adopters, $m(t)$, is a function of time. However, in most empirical contexts, a constant ceiling is considered a good approximation.

3.2 The Bass Framework

As pointed out in Section 3.1 - and as illustrated in Figure 2 - the Bass model is essentially a mixed-influence model that allows for both innovators and imitators. Equation (4) is a first-order

[6] Some early researchers have described this effect as social pressure, rather than a word-of-mouth (i.e. communication) effect (e.g. Bass, 1969).

differential equation with three parameters - the (constant) ceiling of potential adopters, m, the coefficient of innovation, p, and the coefficient of imitation, q.

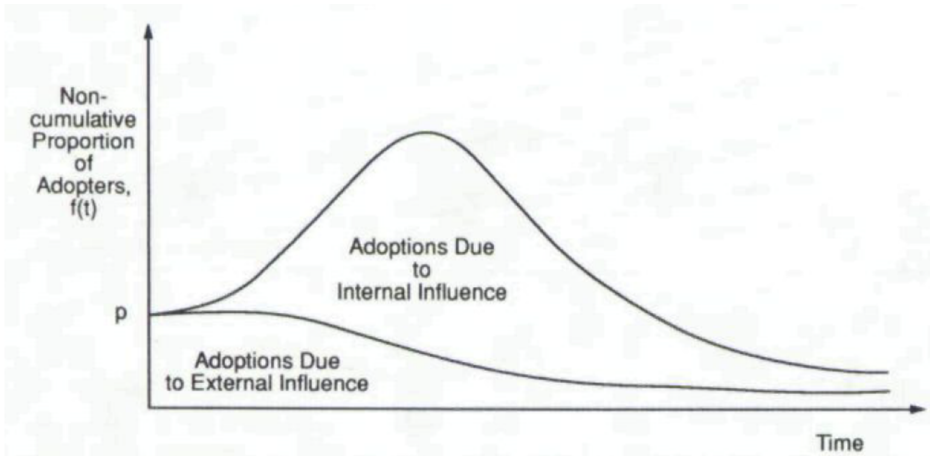

Figure 2: Interplay of Innovation and Imitation in the Bass Model (Mahajan et al., 1990).

Through integration, the solution (Kijek and Kijek, 2010) for the cumulative number of adopters at time t is:

$$N(t) = \frac{m - \frac{p \times (m - N(0))}{p + \frac{q}{m} \times N(0)} \times e^{-(p+q)t}}{1 + \frac{\frac{q}{m} \times (m - N(0))}{p + \frac{q}{m} \times N(0)} \times e^{-(p+q)t}} \tag{5}$$

Given information about the cumulative number of adopters at time zero, $N(0)$, the diffusion curve is fully described by the three Bass parameters, p, q and m.

Much of the early research in the field was concerned with parameter estimation for the Bass model. The four most popular procedures are described by Mahajan et al. (1986):

- ***Ordinary Least Squares (OLS):***

  Bass (1969) used a discrete version of Equation (4):

$$X(i) = p \times m + (q - p) \times N(t_{i-1}) - \frac{q}{m} \times (N(t_{i-1}))^2 \tag{6}$$

  where the newly introduced $X(i)$ represents the incremental increase (i.e. new adoptions) in interval i, so $X(i) = N(t_i) - N(t_{i-1})$.

Equation (6) can be re-written as:

$$X(i) = \alpha_1 + \alpha_2 \times N(t_{i-1}) - \alpha_3 \times (N(t_{i-1}))^2 \tag{7}$$

where $\alpha_1 = p \times m$; $\alpha_2 = q - p$; and $\alpha_3 = q/m$.

Equation (7) can be readily estimated using OLS to get linear least squares estimates for α_1, α_2, and α_3, which can be used to solve for the estimators $\hat{p}$, $\hat{q}$, and $\hat{m}$:

$$\hat{p} = \frac{-\hat{\alpha}_2 + \sqrt{\hat{\alpha}_2{}^2 - 4 \times \hat{\alpha}_1 \times \hat{\alpha}_3}}{2}$$

$$\hat{q} = \frac{\hat{\alpha}_2 + \sqrt{\hat{\alpha}_2{}^2 - 4 \times \hat{\alpha}_1 \times \hat{\alpha}_3}}{2}$$

$$\hat{m} = \frac{-\hat{\alpha}_2 - \sqrt{\hat{\alpha}_2{}^2 - 4 \times \hat{\alpha}_1 \times \hat{\alpha}_3}}{2 \times \hat{\alpha}_3}$$

While OLS estimation is appealing due to its simplicity, it is subject to three problems: First, if estimated with few data points, the OLS estimates are often unstable, and may have wrong signs, due to the highly likely collinearity between the regressors, $N(t_{i-1})$ and $(N(t_{i-1}))^2$. Second, the lack of standard errors for the estimates makes inference based on the model results hard. Finally, the discretization of the continuous process in Bass (1969) in order to use OLS regression introduces a time-interval bias (Boswijk and Franses, 2005; Kijek and Kijek, 2010).

- ***Maximum Likelihood (ML):***

The ML estimation procedure, introduced by Schmittlein and Mahajan (1982), attempts to resolve the problems with OLS estimation. Using, instead of the discrete specification from Bass (1969), the distribution of adoption times as the starting point:

$$F(t) = \frac{c \times (1 - e^{-bt})}{1 + a \times e^{-bt}} \tag{8}$$

where $F(t)$ is the unconditional probability for adoption by time t, c is the probability of eventual adoption, $a = q/p$, and $b = p + q$.

Using the relationship in Equation (8), the likelihood function of the diffusion process can be written as:

$$L(a;b;c;x_i) = (1 - F(t_{T-1}))^{x_T} \prod_{i=1}^{T-1} (F(t_i) - F(t_{i-1}))^{x_i} \tag{9}$$

where x_i denotes the number of individuals who adopt in the interval $i = (t_i, t_{i-1})$. The ML estimators for the parameters a, b and c were then obtained by Schmittlein and Mahajan (1982) using the Hooke-Jeeves accelerated search pattern, as no explicit formulae exist for the parameters which maximize the likelihood function.

Once ML estimates $\hat{a}, \hat{b}$, and $\hat{c}$ are obtained, it is possible to get estimates for p, q, and m through the following relationships:

$$\hat{p} = \frac{\hat{b}}{\hat{a}+1}; \qquad \hat{q} = \frac{\hat{a} \times \hat{b}}{\hat{a}+1}; \qquad \hat{m} = \hat{c} \times M$$

where M denotes the number of potential adopters in the population.[7]

The ML procedure overcomes some of the shortcomings of OLS estimation. The main advantages of the method are that first, we can gain approximate standard errors of the parameter estimates and second, the obtained estimates are more stable and always have the correct sign (i.e. p, q, and m are all non-negative). However, it relies heavily on regularity conditions and is therefore not necessarily suitable with small sample sizes. In particular the assumption that individual adoption times are independent draws seems highly questionable, given that imitation is an important driver of the diffusion path.

- ***Non-linear Least Squares (NLS):***

The NLS approach due to Srinivason and Mason (1986) is an attempt to overcome the shortcomings of the ML approach, most notably the fact that the ML approach underestimates

[7] Note that the introduction of the term c, the probability of eventual adoption, means that Schmittlein and Mahajan (1982) allow for the possibility that not all potential adopters will eventually adopt – other than the original Bass (1969) specification.

standard errors, as it takes into account only sampling errors, but no other sources of error (such as the effects of excluded marketing variables). The NLS approach also seeks to produce valid standard errors for the parameter estimates.

The authors use the following specification:

$$X(i) = N(t_i) - N(t_{i-1}) + \varepsilon \tag{10}$$

where ε is an additive error term that accommodates sampling error, but also all other sources of error. As pointed out by Kijek and Kijek (2010), the parameters p, q, and m (and the respective standard errors) can be directly estimated by non-linear least squares regression using the relationship from Equation (5):

$$X(i) = \frac{m - \frac{p \times (m - N(0))}{p + \frac{q}{m} \times N(0)} \times e^{-(p+q)t_i}}{1 + \frac{\frac{q}{m} \times (m - N(0))}{p + \frac{q}{m} \times N(0)} \times e^{-(p+q)t_i}} - \frac{m - \frac{p \times (m - N(0))}{p + \frac{q}{m} \times N(0)} \times e^{-(p+q)t_{i-1}}}{1 + \frac{\frac{q}{m} \times (m - N(0))}{p + \frac{q}{m} \times N(0)} \times e^{-(p+q)t_{i-1}}} + \varepsilon_i \tag{11}$$

NLS estimates are held to perform at least as well as the ML approach in terms of fit and forecasting accuracy (Srinivason and Mason, 1986). The procedure is able to produce more reliable standard errors and overcomes the time aggregation bias of the OLS approach. Hence NLS is the most commonly used method for the estimation of the Bass parameters.

NLS estimation requires starting values for the model parameters. MLEs are the recommended starting values (Srinivason and Mason, 1986), but results do not differ much when OLS estimates are used as starting values.

- ***Algebraic Estimation (AE):***

The Algebraic Estimation approach (Mahajan and Sharma, 1986) has the virtue of simplicity in obtaining reasonably good estimates of the Bass parameters. It is usually inferior to NLS estimates in terms of fit and forecasting accuracy, but can be used to provide reasonable starting values for the NLS procedure.

The reasoning of Mahajan and Sharma (1986) is that the sophisticated ML and NLS approaches depend heavily on search algorithms and thus are sensitive to the quality of starting values. Poor starting values might lead to slow convergence or even failure to reach the global optimum.[8] Their aim was a simple procedure that requires only knowledge of the inflection point, which could be obtained using data, the experience of analogous products, or even expert opinions (Mahajan and Sharma, 1986).

They argue that knowledge of the inflection point $t^\star$ provides knowledge about $F^\star$, the cumulative fraction of adopters at time $t^\star$, and $f^\star$, the incremental fraction of adopters at time $t^\star$. Given $t^\star$, $F^\star$ and $f^\star$, the algebraic estimation procedure consists of solving a system of simultaneous equations:

$$t^\star = -\frac{1}{p+q} \times ln(\frac{p}{q}) \tag{12}$$

$$F^\star = \frac{N^\star}{M} = \frac{1}{2} - \frac{p}{2q} \tag{13}$$

$$f^\star = \frac{n^\star}{m} = \frac{q}{4} + \frac{p}{2} - \frac{p^2}{4q} \tag{14}$$

The values of $t^\star$, $N^\star$ and $n^\star$, can be used to solve for p, q, and m by reformulating the system:

$$p = \frac{n^\star \times (m - 2N^\star)}{(m - N^\star)^2} \tag{15}$$

$$q = \frac{n^\star \times m}{(m - N^\star)^2} \tag{16}$$

$$t^\star = \frac{m - N^\star}{2n^\star} \times ln(\frac{m}{m - 2N^\star}) \tag{17}$$

Equation (17) can be used to find m either numerically or by trial and error, and once m is known, p and q can be obtained from Equations (15) & (16) (Kijek and Kijek, 2010).

The algebraic estimation procedure is not to be regarded an alternative to NLS estimation, not the least because it cannot produce standard errors for the parameter estimates. However, given its simplicity it is a useful source of starting parameter values for NLS estimation.

[8]For a good mathematical analysis on the conditions that have to be fulfilled to guarantee the existence of NLS estimates, see Jukic (2011, 2013a, 2013b).

It is also appealing that the AE approach always produces estimates with the correct sign (Mahajan and Sharma, 1986).

The Bass model – though originally developed in a marketing context to model the diffusion of innovative products – has become the most commonly employed diffusion model in general. Myriad studies have confirmed its general applicability in modelling technology diffusion (e.g. Dodds, 1973; Chu and Pan, 2008; Michalakelis et al, 2008; Wu and Chu, 2010; Naseri and Elliott, 2013). We choose it as the canonical model against which we compare and benchmark our time-series models, as the Bass model more often than not outperforms the other two commonly used growth models in diffusion studies – the Gompertz and the Logistic growth model, respectively (Naseri and Elliott, 2013).[9] This is particularly true when the upper limit of the market is unknown (Young, 1993). As our objective is to examine the universal suitability of the state-space modelling approach to model diffusion curves, we cannot assume that the upper limit is generally known for any given diffusion process. This further supports the choice of the Bass model as the base model. For obtaining parameters for the Bass model, we use the most frequently used estimation methods – OLS, and NLS with the OLS estimates as starting values. It is worth mentioning that there are several other interesting estimation approaches that we do not explore in detail in this review section – in particular the discrete Bass model (Satoh, 2001) designed to overcome the time aggregation bias of the OLS estimation for the standard Bass model, and the moving least squares method (Scitovski and Meler, 2002), which can be used not only to obtain starting values for the NLS or ML estimation, but can also provide reasonable parameter estimates when no close-form solution of the diffusion model exists (i.e. when it is not possible to use NLS or ML estimation).

4 Time-Series Models

4.1 Box-Jenkins ARIMA Framewok

The ARIMA framework (Box and Jenkins, 1970), is a purely empirical univariate modelling approach that is suitable when the objective is to forecast a (dependent) variable using its own past values. Fitting an ARIMA(p,d,q) model involves two steps:

- Non-stationary features of the time series – such as trend and seasonal components, which are nuisance parameters in the Box-Jenkins framework – are removed from the observed

[9] Despite the findings of Naseri and Elliott (2013), it should be noted that some earlier studies that analysed the performance of the Bass, Logistic, and Gompertz growth models for fitting diffusion curves came up with inconclusive results (e.g. Young and Ord, 1989; Young, 1993).

series by differencing. For the resulting series, the number of differences required to render it stationary is the order of integration, d.

- In modelling the stationary series, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) help to identify the combination of auto-regressive (AR) and moving-average (MA) components that are together able to describe its dynamics satisfactorily. The basic structure of an ARMA(p,q) model (where p denotes the autoregressive order and q denotes the moving average order) can be expressed as:

$$\Phi(L)X_t = \Theta(L)e_t \tag{18}$$

  where X_t is the time series data; e_t is White Noise; $\Phi(L) = 1 - \phi_1(L) - \phi_2 L^2 - \cdots - \phi_p L^p$; $\Theta(L) = 1 - \theta_1(L) - \theta_2 L^2 - \cdots - \theta_q L^q$; and L the lag operator.

  The ARMA(p,q) model is usually fitted using Maximum Likelihood estimation.

The ARIMA approach is most appropriate when the emphasis is on forecasting and when potentially relevant independent variables cannot be forecast effectively (Studenmund, 2006). ARIMA models generally produce more reliable short-term forecasts than traditional econometric models (Gujarati, 2004), and the differential equation models we discussed in Section 3.

4.2 State-Space Models[10]

The state-space approach provides a rich family of models for time series data based on the essential idea that behind an observed time series there is an underlying, general, but hidden state process which evolves over time in a way that reflects the structure of the system being observed. The observed series are written as a function of the unobserved state process, which itself is governed by a first-order Markovian transition equation, contaminated by noise. Using a model that consists of the state transition equation and an equation relating observations to the state, the task is to estimate the hidden state process, which is the feature of interest.

State-space models, which encompass structural time series models in econometrics, treat trend, cycle, and seasonality as parameters of interest, rather than nuisance parameters as in the ARIMA

[10]Note that only linear Gaussian state-space models are reviewed here. For a general review of non-linear non-Gaussian state-space models, see Durbin and Koopman (2001); parameter estimation problems in the non-linear non-Gaussian setting are discussed by e.g. Tanizaki and Mariano (1998), Tanizaki (2001), Andrieu et al. (2005), and Strickland et al. (2008).

framework. The ability to extract the unobserved dynamic features of a process in an efficient and interpretable way makes state-space models a potentially powerful framework for diffusion research.

State-space models have been discussed in detail in the literature. The purpose of our project is to apply the methodology in a new context and to evaluate its suitability (rather than to develop or extend the state-space framework itself). A detailed review of the methodology would far surpass the scope of this paper. We only present the general idea behind the state-space framework and its advantages and disadvantages in comparison to ARIMA modelling as a prelude to our empirical analysis. For a good review of the method, the interested reader is referred to the following books and papers: early books (mainly engineering research) include Jazwinski, 1970, Sage and Melsa (1971), and Anderson and Moore (1979). In the statistical/econometric context, pioneering work is due to Harvey (1989). Other work that treats the method extensively includes Kitagawa and Gersch (1996), as well as Durbin and Koopman (2001). The parameter estimation problem is discussed in detail by Harvey and Peters (1990), and Davis and Rodriguez-Yam (2005), while much of the more recent research focuses on Bayesian parameter estimation (e.g. West and Harrison, 1997; Frühwirth-Schnatter, 2006). Even though state-space models have been extensively studied in the literature, the implementation of the framework in non-specialist statistical software and packages is limited. This makes careful reviews of practical application of the method valuable – these include Teyssiere (2005), Petris et al. (2009), Mendelsohn (2011), and Petris and Petrone (2011).

The state of the system is defined as the minimum set of information from the past and present such that the future behavior of the system can be fully described by the knowledge of the present state and future input (Wei, 1990). This description of the state-space framework has an important implication: State-space models are based on the Markov property, i.e. future behavior of the system is independent of its past behavior, given the present state. A univariate time series can then be described in its state space form as follows (Harvey and Peters, 1990):

$$y_t = z_t' \alpha_t + \varepsilon_t \tag{19}$$

$$\alpha_t = T_t \alpha_{t-1} + \xi_t \tag{20}$$

where Equation (19) is the observation (or measurement) equation and Equation (20) is the state equation (or transition equation). y_t are the observed values, α_t is the state vector of dimension $m \times 1$, z_t is an $m \times 1$ fixed vector (that is usually called the observation or design vector), T_t is an $m \times m$ fixed transition matrix, and ε_t (a scalar) and ξ_t (a vector of dimension $m \times 1$) are the disturbances which are distributed normally and are independent of each other. The aim of the analyst is to produce an estimator of the underlying unobserved signals $\alpha^t = (\alpha_1, ..., \alpha_t)$ given the observed time series $y^s = (y_1, ..., y_s)$. If $t > s$, the problem is one of forecasting. If $t = s$, it is a filtering problem and for $t < s$, it is a smoothing problem.[11]

The general state-space representation shows the flexibility of the approach. The advantage over the ARIMA approach is obvious: ARIMA models can always be written in state-space form and is thus a special case of the state-space representation (Wei, 1990). Cowpertwait and Metcalfe (2009) argue that the main merit of state-space models is the ability of their parameters to adapt over time and to do so quickly. This is an advantage over common time-series smoothing techniques like the Holt-Winters framework[12]).

Other advantages of the state-space framework include (Harvey, 1989; Durbin and Koopman, 2001):

- It can handle shifts, breaks, and time-varying parameters of a static model well (by making them dynamic states),
- It can handle missing data relatively well (by skipping the updating step and just executing the transition step in the Kalman filter estimation),
- The model consists of directly interpretable unobservable dynamic components which can be estimated,
- Explanatory variables can be introduced in the model easily,
- It does not require stationarity of the time series, and so obviates the need for data transformation.

In summary, it can be said that the state-space approach provides a structural framework for the decomposition of time series, enabling simultaneous analysis of all the salient unobserved dynamics in the time series data. The fact that non-stationary features of the data are treated as parameters of interest (rather than nuisance parameters as in the Box-Jenkins approach) makes state-space modelling a much richer framework whenever the task is not merely a forecasting problem.

[11] For a discussion of smoothing, filtering, and forecasting structural time series, see Harvey and Shephard, 1993.

[12] For an overview of exponentially weighted moving average models (including the Holt-Winters framework) in a structural time series context, see Harvey, 1989, Chapter 2.

5 Empirical Analysis

5.1 The Data

Data utilized in this study are drawn from Manuelli and Seshadri (2014). The data displays the diffusion of tractors on U.S. farms from their introduction in 1910 to 1960. Taking a first look at the series, it can be seen that the diffusion process coincides with general diffusion theory overall. After a slow diffusion rate from 1910 - 1940, the number of new tractors is sharply rising until 1950 (with some fluctuations) and shows signs of saturation after 1950. This translates into the characteristic S-shaped curve when considering the cumulative number of tractors in use. A notable feature of the series is the considerable dip in the early 1930s, when the increasing trend of the series is broken abruptly and the cumulative number of tractors in use stagnates for four to five years. We will take a closer look at this potential saddle in Section 9 and will now focus on fitting the series with the models discussed in the previous sections.

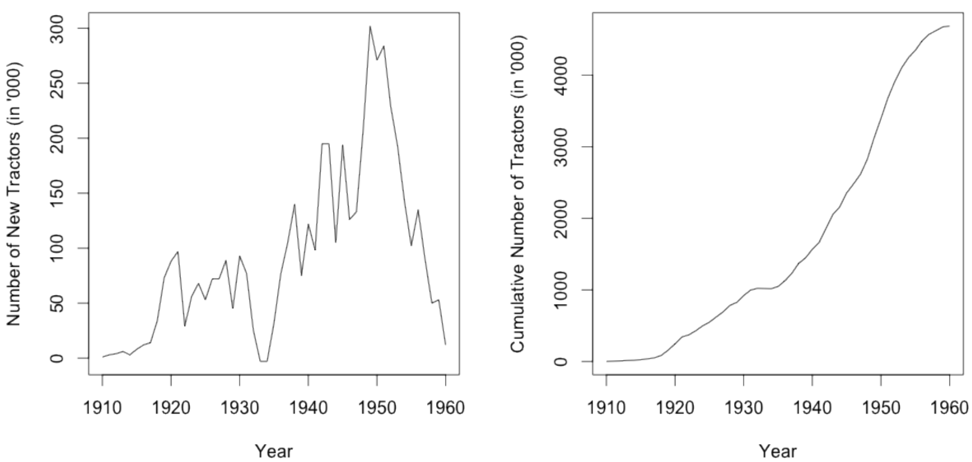

Figure 3: (a) Number of New Tractors (in Thousands), 1910 - 1960; (b) Cumulative Number of Tractors in Use (in Thousands), 1910 - 1960.

5.2 Classical Diffusion Models:

- ***Bass Framework:***

  As discussed in Section 3, the most popular estimation methods for the Bass model are the OLS method by Bass (1969) and the NLS method by Srinivason and Mason (1986). Therefore, we estimated both models. The OLS specification is equivalent with the one presented in Bass (1969) – recall Equation (7):

$$Bass(OLS) : X(i) = \alpha_1 + \alpha_2 \times N(t_{i-1}) - \alpha_3 \times (N(t_{i-1}))^2 \qquad \text{(7 revisited)}$$

In our case, $X(i)$ is the incremental change in the number of tractors from year t_{i-1} to t_i, and $N(t_{i-1})$ is the cumulative number of tractors in use up to year t_{i-1}.

The model was fitted using R statistical software (R Development Core Team, 2012).[13] The linear regression model returned the following results (standard errors of the estimates are presented in brackets):

$$Bass(OLS) : X(i) = \underset{(13.59)}{-0.4967} + \underset{(0.018)^{***}}{0.1278} \times N(t_{i-1}) - \underset{(3.916 \times 10^{-6})^{***}}{(-2.309 \times 10^{-5})} \times (N(t_{i-1}))^2$$

From this model we can extract the estimates for the Bass parameters, i.e. p̂, q̂, and m̂:[14]

$$Bass(OLS) : \hat{p} = -8.9803 \times 10^{-5}$$

$$Bass(OLS) : \hat{q} = 0.1277$$

$$Bass(OLS) : \hat{m} = 5531.365$$

Overall, the model explains the variability in the number of new tractors in use quite well with an adjusted R^2 of 55.73 %. However, looking at the values of the estimates, one of the drawbacks of the OLS estimation method becomes apparent quickly - the estimate for the

[13] All following models in this section were also fitted in R, unless stated otherwise.

[14] $\hat{p} = \frac{-0.1278+\sqrt{0.1278^2-4\times-0.4967\times-2.309\times10^{-5}}}{2} = -8.9803\times10^{-5}$

$\hat{q} = \frac{0.1278+\sqrt{0.1278^2-4\times-0.4967\times-2.309\times10^{-5}}}{2} = 0.1277$

$\hat{m} = \frac{-0.1278-\sqrt{0.1278^2-4\times-0.4967\times-2.309\times10^{-5}}}{2\times-2.309\times10^{-5}} = 5531.365.$

coefficient of innovation, $\hat{p}$, is negative, which makes interpretation hard. Other than that the parameter estimates seem reasonable: $\hat{p}$ is close to 0, which coincides with the empirical findings of Jeuland (1995), as well as Sultan et al. (1996). The estimate for the coefficient of imitation, $\hat{q}$, seems fairly low compared to the typical range of 0.3 - 0.5 that was stated by Mahajan et al. (1995). This could indicate that the word-of-mouth effect was rather weak among the group of potential adopters for this specific technological innovation, which somewhat coincides with the findings of Manuelli and Seshadri (2014), who present evidence that in the market introduction phase, farmers were slow to adopt tractors, even though they were the cost-minimizing technology. The estimate for the upper limit of the market, $\hat{m}$, seems quite reasonable, given that at the end of the series signs of a saturated market can already be seen – as the cumulative number of tractors in use in 1960 was 4,685,000, this suggests that about 84.52 % of potential adopters had already adopted tractors on their farms.

Figure 4 plots the fitted values versus the actual values for both the incremental and the cumulative series. It can be seen in Figure 4(a) that the model captures the major trends in the data quite well, but is not able to adjust to short-term fluctuations or shocks. This is rather unsurprising, as the only explanatory variables to explain the new number of tractors in a given year are transformations of the lagged cumulative series. One notable feature is that (again due to the lagged cumulative series as explanatory variable) the stagnating phase in the 1930s is captured by the model to some extent. However, even though the stagnation is captured, the strong dip in the number of tractors is not fitted very well and the model seems unable to capture the increasing adoption rate post-1940.

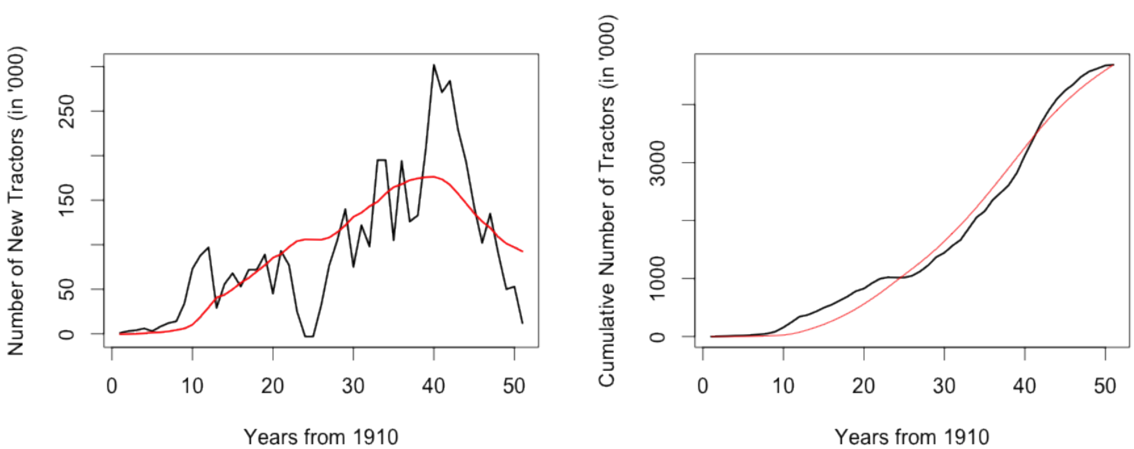

Figure 4: (a) Actual vs. Fitted (Bass OLS), Incremental Series; (b) Actual vs. Fitted (Bass OLS), Cumulative Series.

The standardized residuals of the model (Figure 5) show severe problems with the OLS assumptions. A first visual inspection already makes it obvious that the residuals do not resemble White Noise, as they show a strong systematic pattern. The problem with the assumption of independence is confirmed by the ACF that shows highly significant values at Lag 1 ($r_1 = 0.637$) and Lag 2 ($r_2 = 0.442$) – the 95 % critical values of the ACF are at $\pm 2/\sqrt{51} = \pm 0.28$. This result was also confirmed using the Ljung-Box Q statistic that tests for joint independence up to a certain lag, which showed highly significant values for all tested lags (up to 30).[15]

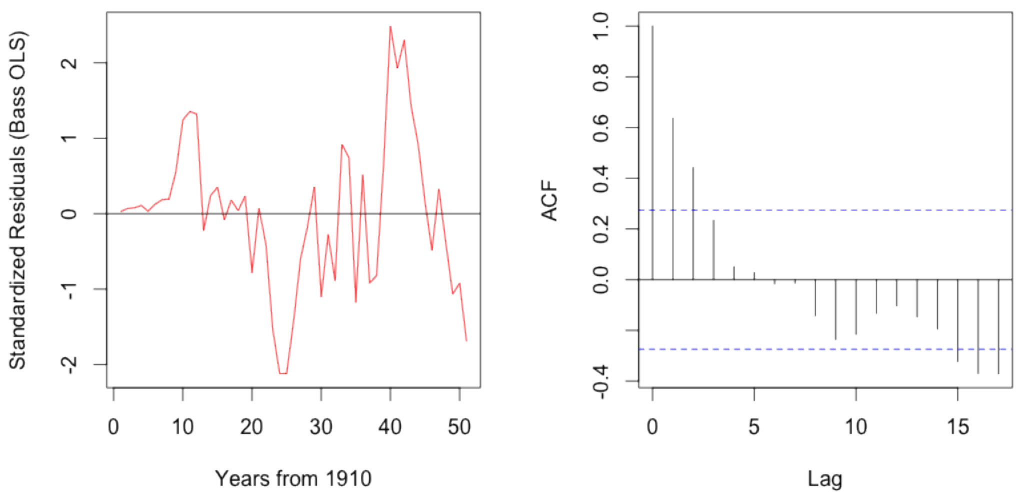

Figure 5: (a) Standardized Residuals (Bass OLS); (b) Autocorrelation Function (Bass OLS).

There also are severe issues with the assumption of homoscedasticity of the residuals, which is confirmed by a Breusch-Pagan test statistic of 7.17, which corresponds to a p-value of 2.78 % and therefore suggests that the residuals suffer from heteroscedasticity (at the 95 % confidence level).

The assumption of normality was checked by a visual inspection of the histogram of the studentized residuals, as well as a normal quantile-quantile plot of the studentized residuals (Figure 6). The plots suggest good approximation of normality – this is confirmed by a Jarque-Bera test with a test statistic of 0.31, which corresponds to a p-value of 85.82 % and consequently does not reject normality at the commonly used significance levels.

[15]The Ljung-Box statistics for the Bass OLS residuals are not presented in the report, but can be provided on request.

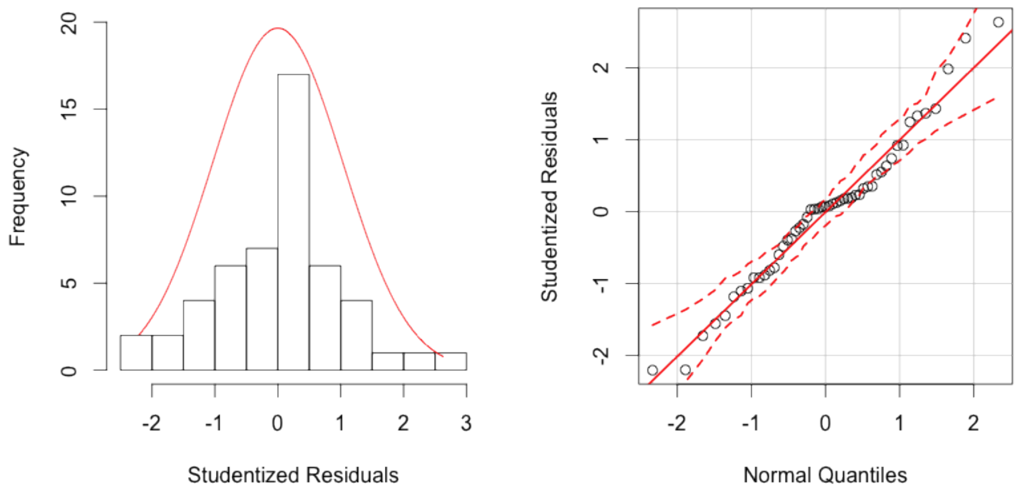

Figure 6: (a) Histogram of the Studentized Residuals (Bass OLS); (b) Normal QQ-Plot of the Studentized Residuals (Bass OLS).

Overall, the residual diagnostics show clear problems of the Bass OLS model – in particular the dependence structure (and to some extent the heteroscedasticity in the residuals) is a strong reason for concern.

For NLS estimation, we used the specification from Cowpertwait and Metcalfe (2009):[16]

$$Bass(NLS): S(t) = \frac{m \times (p+q)^2 \times e^{-(p+q)t}}{p \times (1+\frac{q}{p} \times e^{-(p+q)t})^2} \tag{21}$$

The non-linear least squares estimation was carried out using the Gauss-Newton algorithm (Björk, 1996), with OLS estimates as starting values ($\hat{p}_{start} = -8.9803 \times 10^{-5}$, $\hat{q}_{start} = 0.1277$, and $\hat{m}_{start} = 5,531.365$). The algorithm converges after 18 iterations and reduces the residual sum of squares to 112,678.5 (from 5,081,378 for the model with the OLS starting values).

The model returned the following estimates:

– $\hat{p}$ = 1.0597×10^{-4} (s.e. = 7.546×10^{-5}); the value is still close to 0, but the incorrect sign of the OLS estimate of the coefficient of innovation has been reversed.

[16]For the derivation of the relationship, see Cowpertwait and Metcalfe (2009), p. 52.

– $\hat{q} = 0.1912$ (s.e. $= 0.021^{***}$); while this is lower than the typical values for the coefficient of imitation reported by Mahajan et al. (1995), the NLS model suggests a stronger word-of-mouth effect than the OLS model. The probable reason can be seen in the model fit in Figure 5: The NLS model captured the increased adoption rate in the 1940s better.

– $\hat{m} = 4,472.857$ (s.e. $= 385.90^{***}$); this estimate for m (the upper limit of the market) is not plausible, being lower than the actual cumulative number of tractors in use in 1960.

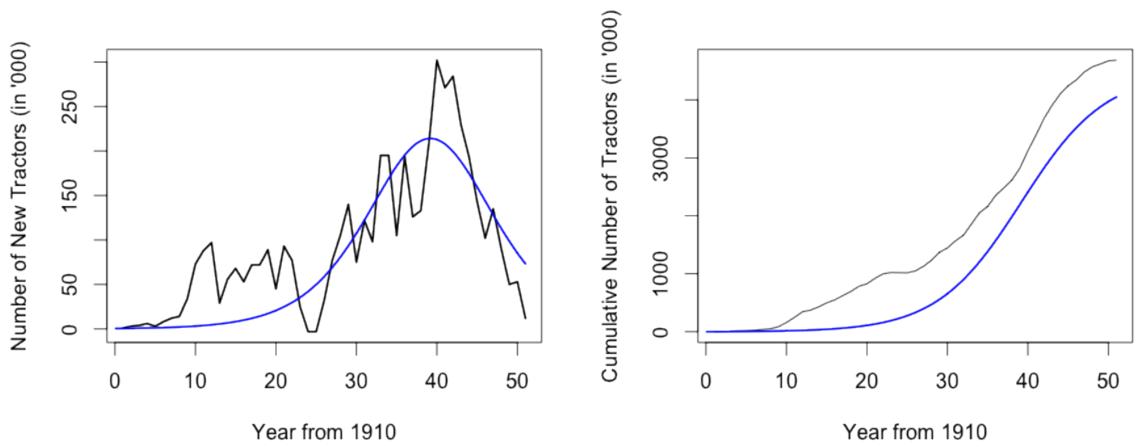

Figure 7: (a) Actual vs. Fitted (Bass NLS), Incremental Series; (b) Actual vs. Fitted (Bass NLS), Cumulative Series.

The model fit of NLS estimation is shown in Figure 7 for both the incremental and the cumulative series. Figure 7(a) shows that the change in the adoption rate was captured quite well by the NLS model. The adoption rate was moderate up to approximately 1940, and then changed to have a steep slope. Like the OLS model, the NLS model also captures the inflection point around 1950. Overall, the NLS model takes into account the extreme observations (the sharp drop in the 1930s and the maximum around 1950) much better than the OLS model. An important consequence of fitting the extreme points is the underestimation of the number of new tractors in the early phase of the observation period, resulting in the underestimation of the entire cumulative curve as seen in Figure 7(b). The cumulative fit for the later phase of the observation period appears to be good, with the estimated curve amounting to a parallel shift of the observed series. The results indicate that a break that disrupts the classical shape of the diffusion process cannot lead to poor estimates from this family of models.

The residuals of the NLS model show the same problems as those from the OLS model and are therefore not presented in this report (available on request).

- ***Logistic Model:***

The logistic model is based on the logistic differential equation:

$$LogisticGrowth: \frac{dN(t)}{dt} = r \times N(t) \times (1 - \frac{N(t)}{m}), \quad N(0) = N_0 \tag{22}$$

where N_0 is the initial density of the population, r is the growth rate of the population and m is the maximum potential population density.

This differential equation can be solved analytically for $N(t)$ by separation of variables (Sterman, 2000)[17]:

$$LogisticGrowth: N(t) = \frac{m}{1 + \frac{m - N_0}{N_0} \times e^{-rt}} \tag{23}$$

The explained variable in this framework is the cumulative number of tractors in use and the model parameters are m, the upper limit of the market potential, N_0, the (cumulative) number of adoptions at time 0, and r, the intrinsic growth rate of the process. Again, we used non-linear least squares to estimate the model. As starting values for the parameters, we used the OLS estimate for the market potential for m (i.e. 5,531.365), 0.1 for the initial number of tractors (Note: using 0 led to computational issues), and 0.5 for the intrinsic growth rate, r.

The algorithm converged after 23 iterations and returned the following estimates: 83.04 (s.e. = 11.78***) for $\hat{N}_0$, 0.1061 (s.e. = 0.005***) for the growth rate, $\hat{r}$, and 6,762 (s.e. = 397.40***) for the upper limit of the market potential, $\hat{m}$. It is interesting to note that the logistic model generated a considerably higher market potential estimate than the Bass model. The growth rate is plausible, but the estimate for the initial value is far too high, as the data starts with the introduction of the tractor in the U.S. market, i.e. at a number of 0.

Figure 8 shows the model fit of the logistic model for both the incremental and the cumulative series. Like the Bass model estimates, the logistic growth model also captures the major trends. Overall, it seems closer to the Bass OLS model, as it does not capture the change in slope as sharply as the Bass NLS model does, and it does not seek to fit the extreme points as much. The dip and the stagnation in the series are not captured well by the logistic model at all compared to the Bass framework, whereas the OLS version captured the stagnation effect and the NLS version fit the level of the dip to some extent. It is not surprising that this phase

[17] The steps of the analytical solution of the logistic differential equation are presented in Appendix A.

is not shown by the model that well, as the logistic model fits the cumulative series, while the Bass model uses the incremental series as dependent variable. From the cumulative plot it looks as if the logistic model has not captured the saturation tendency towards the end of the series very well, which raises questions about the high estimate for the market potential compared to the Bass models.

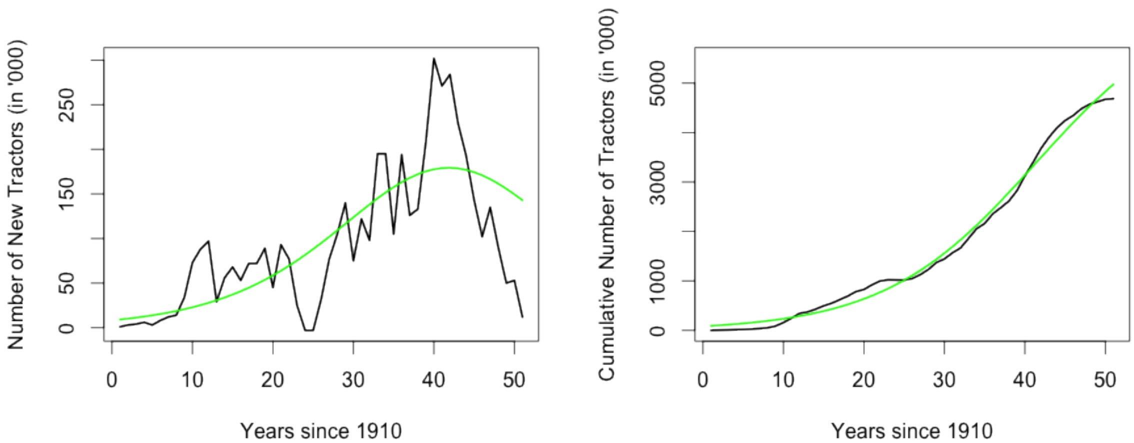

Figure 8: (a) Actual vs. Fitted (Logistic), Incremental Series; (b) Actual vs. Fitted (Logistic), Cumulative Series.

The residuals shows similar signs of dependence and heteroscedasticity as in the Bass OLS case, which is why no residual diagnostics are presented for the logistic growth model in this report (available on request).

Overall, the models based on differential equations are useful in capturing the long-term (stylised) behavior. They give reasonable estimates of the market potential as long as sufficient historical data is available. They can provide some information about the expected growth rates of the market. One obvious shortcoming of the differential equation (hazard) approach is their inability to capture short- and medium-term breaks properly. This leads to inaccurate short-term predictions, which is a key purpose in the marketing diffusion literature. Another issue is that the residuals of all three differential equation-based models siffer from dependence and heteroscedasticity.

- ***ARIMA Model:***

As discussed in Section 4.1, ARIMA models require stationary inputs – i.e. with constant mean and variance and a covariance pattern that only depends on lags, not on time. The tractor series shows a clear upward trend until 1950 and a sharp decline after. Furthermore, the variance appears to increase with time. The ACF shows a slow decay, and the Augmented

Dickey-Fuller test for the unit root null hypothesis, returned an insignificant p-value of 68.96 %. The series needs to be differenced in order to proceed.

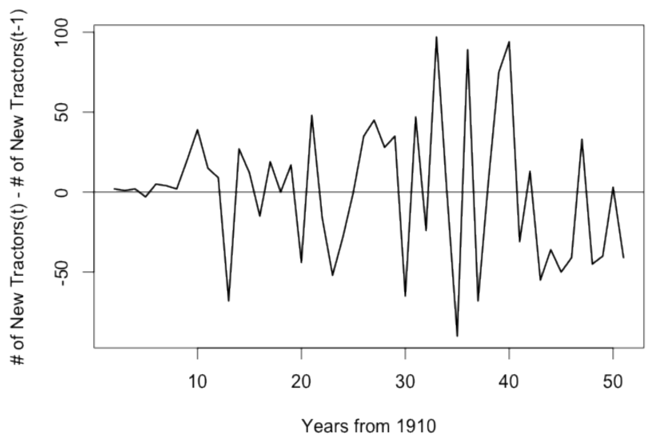

Figure 9: First-Difference of Original Series.

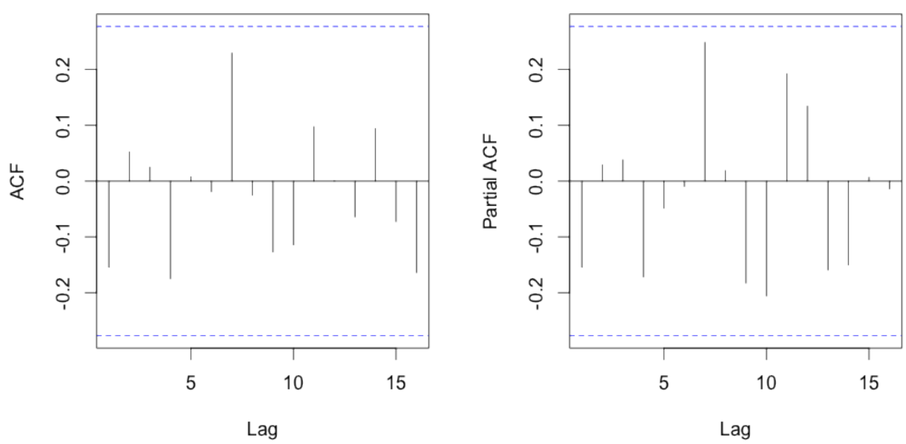

Figure 10: (a) ACF Plot of the First-Differenced Series; (b) PACF Plot of the First-Differenced Series.

Both the time-series plot (Figure 9) and the ACF plot (Figure 10(a)) show no issues with stationarity for the first-differenced series. The Augmented Dickey-Fuller test returned a p-value of 1.98 %, leading us to reject the null hypothesis of a unit root at the 5 % significance level.

We proceed with the first-differenced series and select suitable ARMA orders. The ACF plot in Figure 10(a) does not show any significant lags, indicating no significant MA terms. The same holds true for the PACF plot in Figure 10(b), suggesting that the AR order is zero as well. Consequently, the first model that was fitted was an ARIMA(0,1,0) – a random walk.[18]

The ARIMA(0,1,0) model, estimated using Maximum Likelihood returned an AIC of 518.48. In order to evaluate the model, we applied an overfitting procedure. First, we tried to overfit an AR term, i.e. an ARIMA(1,1,0) model. Although the AR(1) term was significant, the AIC deteriorated to 519.46. Analogously, we tried to fit an ARIMA(0,1,1) model, which also returned a significant MA(1) term, but deteriorated the AIC to 519.55. Hence, as the overfitted models showed no apparent significant improvement, leading us to the random walk model as our final model.

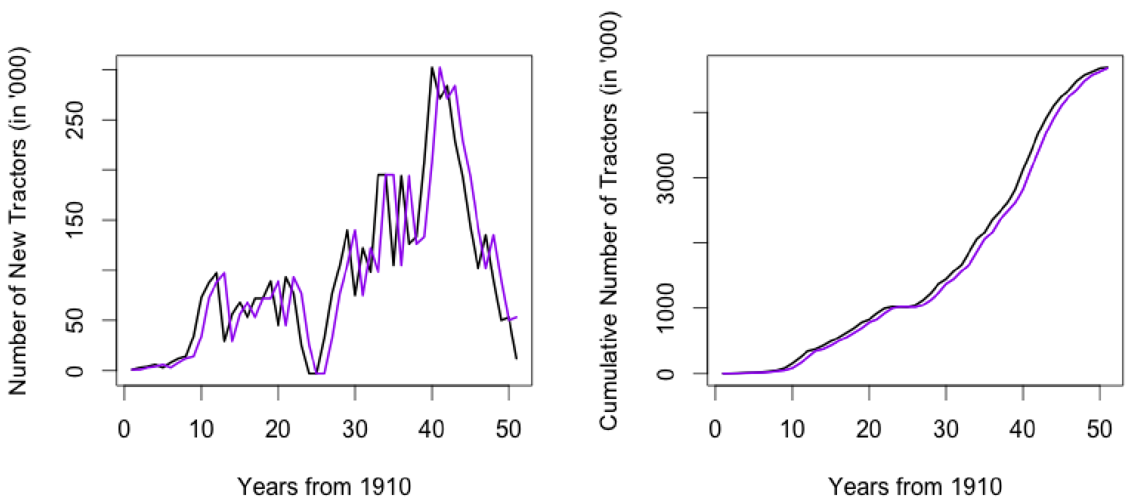

Figure 11: Model Fit of the ARIMA(0,1,0).

The model fit can be seen in Figure 11. The random walk model suggests that the change in the number of new tractors from one year to the next is independent of the number of new tractors in the past year. From the plot it seems to provide a much better fit for the data than the differential equation-based models. This comes at a cost – while we can use the ARIMA model to produce more reliable short-term forecasts (as will be shown in the next subsection), the random walk model has no parameters that can help us understand the characteristics of the long-term diffusion path; in particular the lack of an estimate of the eventual market potential is a major drawback of the ARIMA method in a diffusion context. Of course, as reported in the previous section, the parametric differential equation-based models do not always produce realistic parameter estimates.

[18] As pointed out by Kendall (1973), as well as Harvey and Todd (1983), it is a common problem of ARIMA modelling that for small datasets the correlogram and the partial autocorrelation function are not very informative.

6 State-Space Models

6.1 Estimation & Model Selection

Our objective is to introduce and establish state-space models the more suitable approach to diffusion modelling, compared to extant empirical methods. Therefore we begin with a discussion of the framework and model selection within it, before presenting results from our estimation.

Unlike the case with ARIMA models, the state-space approach does not require stationarity in the series under analysis – we can can proceed to the estimation directly using the incremental series. Our proposed approach for univariate state-space model selection in a diffusion context is presented in Figure 12.

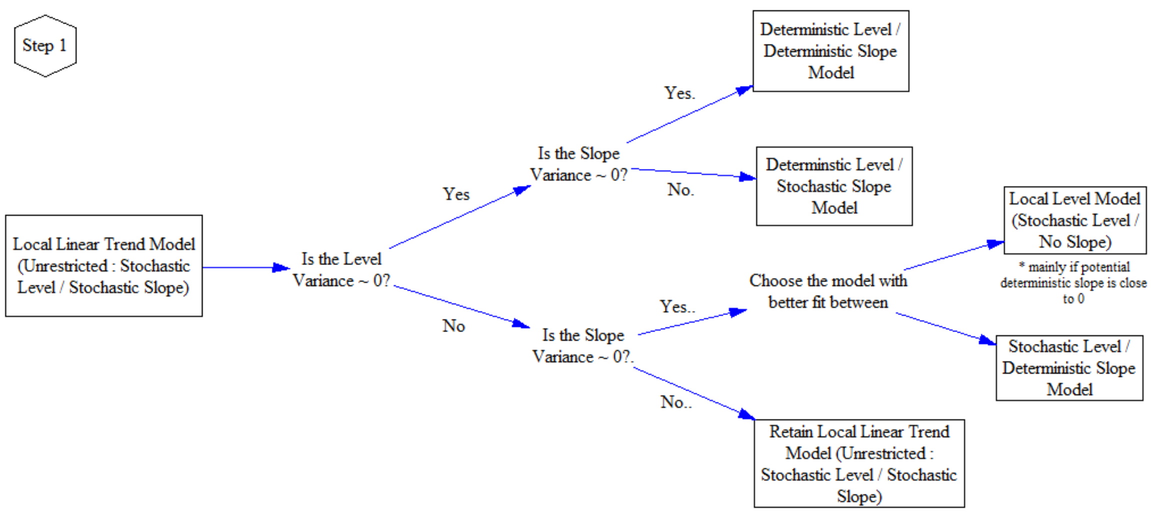

Figure 12: Proposed Approach for State-Space Model Selection in a Diffusion Context (Case Without Interventions or Explanatory Variables).

We begin with a general model specification at the outset, subject to the rule that a cyclical component be included only upon evidence that a cycle is indeed present in the data, as identifiability issues arise otherwise (Harvey, 1989). The general approach is that if the variance of the disturbance pertaining to a state variable (i.e., the variance of the disturbance in the transition equation of the state variable) is found to be close-to-zero, then that is suggestive that the corresponding state component may as well be treated as a deterministic effect, with the advantage of parsimony (Andrews, 1994; Commandeur and Koopman, 2007).

The general/unrestricted model that we apply to the start off the analysis of our annual data series allows for a time-varying level, as well as a time-varying slope – the local linear trend model in the literature:[19]

$$X_t = \mu_t + \varepsilon_t, \ \ \varepsilon_t \sim \ NID(0, \sigma^2_\varepsilon)$$

$$\mu_{t+1} = \mu_t + \nu_t + \xi_t, \ \ \xi_t \sim \ NID(0, \sigma^2_\xi)$$

$$\nu_{t+1} = \nu_t + \zeta_t, \ \ \zeta_t \sim \ NID(0, \sigma^2_\zeta)$$

The first equation is the observation equation, which describes the observed value in terms of the estimated level (state) and an irregular term. The second and third equations describe the transition processes governing the state variables, i.e. the stochastic level, μ_t, and the stochastic slope, ν_t. This linear Gaussian state-space model is straightforward to estimate via Maximum Likelihood using standard statistical software – we use STAMP, a package specifically designed to estimate state-space models (Koopman et al., 2009).[20]

At convergence, the value of the log-likelihood function is -185.543, which corresponds to an AIC of 381.09.[21] The Maximum Likelihood estimate of the variance of the irregular component ($\hat{\sigma}^2_\varepsilon$) is 230.67. The Maximum Likelihood estimates of the state disturbance variances are $\hat{\sigma}^2_\xi = 1{,}338.92$, and $\hat{\sigma}^2_\zeta = 1.51$, respectively. The Maximum Likelihood estimates of the initial values of the level and the slope are $\hat{\mu}_1 = 1.048$ and $\hat{\nu}_1 = 1.556$, respectively.[22] Figure 13 plots actual

[19] In the case of higher-frequency (e.g. quarterly or monthly) data, it might be desirable to include also a seasonality component in the model, in which case the unrestricted starting model should also allow for a time-varying seasonal effect. For an overview of seasonal components in state-space models, see Harvey (1989), Chapter 6.2, or Commandeur and Koopman (2007), Chapter 4. For a discussion of seasonality in diffusion processes, see Guidolin and Guseo (forthcoming).

[20] For structural time series models, the Maximum Likelihood estimation of the hyperparameters can be carried out in the time domain (by putting the model in a state-space form and applying the Kalman filter) or in the frequency domain – for a detailed discussion, see Harvey (1989), Chapters 4.2 & 4.3, or Harvey and Peters (1990). Once the hyperparameters are estimated, the estimates of the state components can be obtained through a smoothing algorithm. All estimates presented in this section are the smoothed estimates, as opposed to the filtered or predicted estimates. For discussion on this, see Harvey (1989), Chapter 3.6 & 4, or Commandeur and Koopman (2007), Chapter 8.4.

[21] In order to compare the goodness of fit of the estimated state-space models, we use the original version of the Akaike Information Criterion: $AIC = -2 \times logL + 2 \times k]$, where $logL$ is the value of the log-likelihood function at convergence, and k is the number of estimated parameter, i.e. the number of diffuse initial values in the state (since diffuse initialisation is used in the estimation process) + the number of disturbance variances. For a detailed review of the use of AIC in a state-space context, as well as an improved information criterion for state-space model selection, see Bengtsson and Cavanaugh (2006).

[22] The initial values were obtained using diffuse initialisation. For a brief discussion of the use of diffuse priors, as opposed to fixed initial state vectors, see e.g. Harvey (1989); Harvey and Peters (1990).

vs. fitted values (top left), the Maximum Likelihood estimates of the local level (top right), and the local slope (bottom left), both with their approximate 95 % confidence intervals, as well as the irregular component (bottom right).

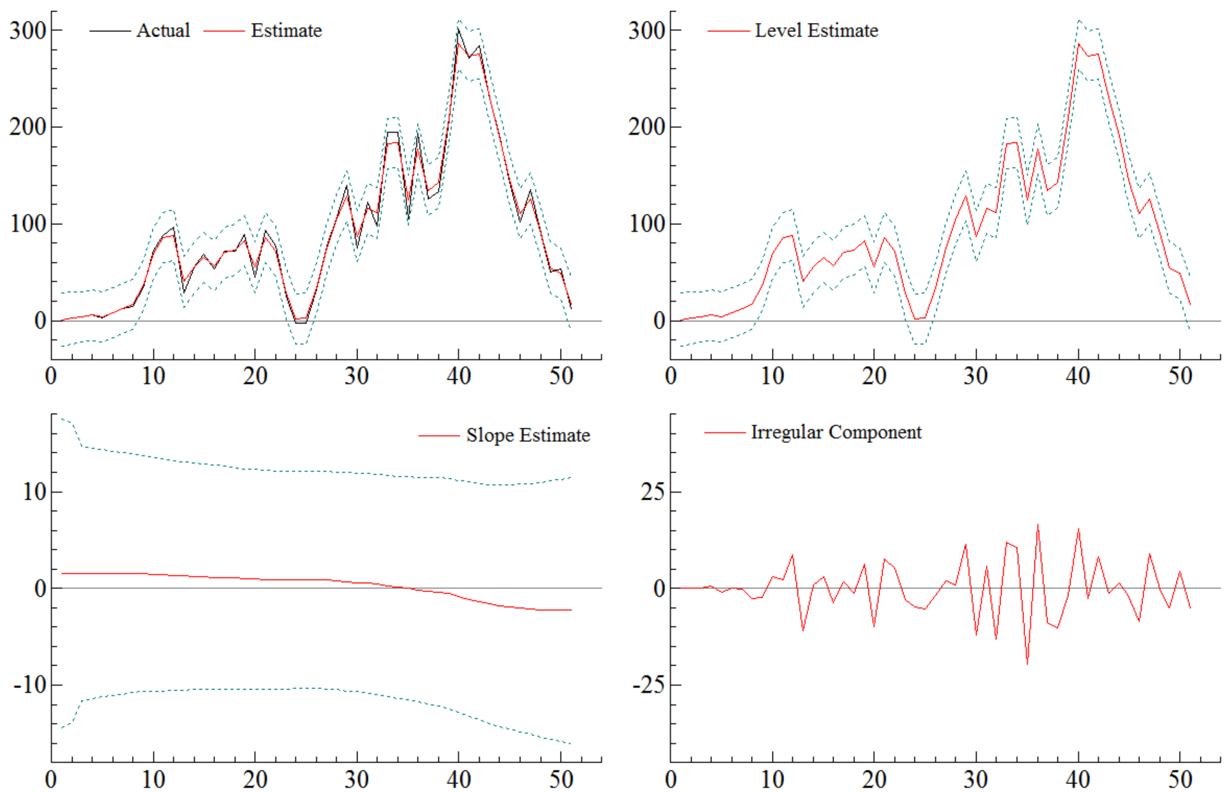

Figure 13: Fit of the Local Linear Trend Model.

We next inspect the estimate for the level variance ($\hat{\sigma}^2_{\xi}$) – the rather high value of 1,338.92 suggests that a deterministic level component is unlikely to be sufficient to describe the data well. The high value of the estimate for the variance of the level disturbance in relation to the irregular component (signal-to-noise ratio of 5.80[23]) suggests that there is considerable discounting of past observations, i.e. the model fit emphasises the most recent values considerably (Harvey, 1989). We next inspect the slope variance ($\hat{\sigma}^2_{\zeta}$). The low value of 1.51 indicates little variation in the local slope estimates, suggesting that a time-varying slope component might not be required. The visual inspection of the local slope component supports this, since the range of the local slope estimates is small compared to the range of the data and there is little fluctuation in the estimates.

Given the high level component variance and low slope component variance, we proceed to estimate two models without time-varying slopes:

[23]The signal-to-noise ratio (often called the q-ratio) is defined as q = $\sigma^2_{\xi}/\sigma^2_{\varepsilon}$ (Harvey and Shephard, 1993).

- A model with stochastic level and deterministic slope (random-walk-plus-drift model (Harvey, 1989)):

$$X_t = \mu_t + \varepsilon_t, \varepsilon_t \sim \ NID(0, \sigma_\varepsilon^2)$$

$$\mu_{t+1} = \mu_t + v_1 + \xi_t, \xi_t \sim \ NID(0, \sigma_\xi^2)$$

  where v_1 denotes the constant slope. This requires the estimation of two diffuse initial values (μ_1 and v_1), but only two disturbance variances, as opposed to three in the local linear trend model. This model returned a log-likelihood of -185.553 at convergence, corresponding to an AIC of 379.10 (and hence provided a slightly better fit than the local linear trend model). The Maximum Likelihood estimate of the variance of the irregular component ($\hat{\sigma}_\varepsilon^2$) is 218.84. The Maximum Likelihood estimate of the level disturbance variance ($\hat{\sigma}_\xi^2$) is 1376.04. The Maximum Likelihood estimates of the initial values of the level and the slope components are $\hat{\mu}_1 = 1.219$, and $\hat{v}_1 = 0.318$, respectively.

  The model fit (plotted in Figure 14) does not seem to differ hugely from the local linear trend model. This is not unexpected, given the similar AIC value and Maximum Likelihood estimates of the hyperparameters.

- Since the Maximum Likelihood estimate of the deterministic slope component in the previous model was close to zero (insignificant at the 95 % level), it might be that a slope component is not required at all in order to describe the diffusion path well. Therefore, we also estimated a model with only a stochastic level component and no slope component – this model is called a local level model or random-walk-plus-noise model (Harvey, 1989):

$$X_t = \mu_t + \varepsilon_t, \varepsilon_t \sim \ NID(0, \sigma_\varepsilon^2)$$

$$\mu_{t+1} = \mu_t + \xi_t, \xi_t \sim \ NID(0, \sigma_\xi^2)$$

  The local level model requires the estimation of two disturbance variances and one diffuse initial value (μ_1) and is therefore the most parsimonious of the three models presented. However, the model does not fit the data as well, with the log-likelihood at -186.699 upon convergence, corresponding to an AIC of 379.40. The Maximum Likelihood estimate of the variance of the irregular component ($\hat{\sigma}^2_{\varepsilon}$) is 250.18. The Maximum Likelihood estimate of the level disturbance variance ($\hat{\sigma}^2_{\xi}$) is $1,285.71$ and the Maximum Likelihood estimate of the initial value of the level ($\hat{\mu}_1$) is 1.311.

The fit of the local level model is presented in Figure 15. It can be seen that a stochastic level (without any slope component) is able to describe the data quite well. The stochastic level term is able to capture the path of the diffusion process over the entire observation period. There is no apparent systematic pattern in the irregular term.

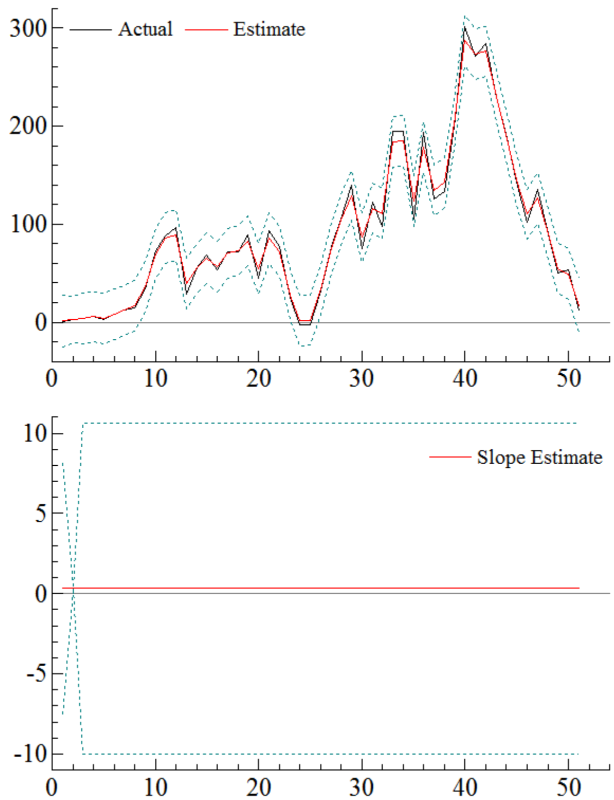

Figure 14: Fit of the Stochastic Level/Deterministic Slope Model.

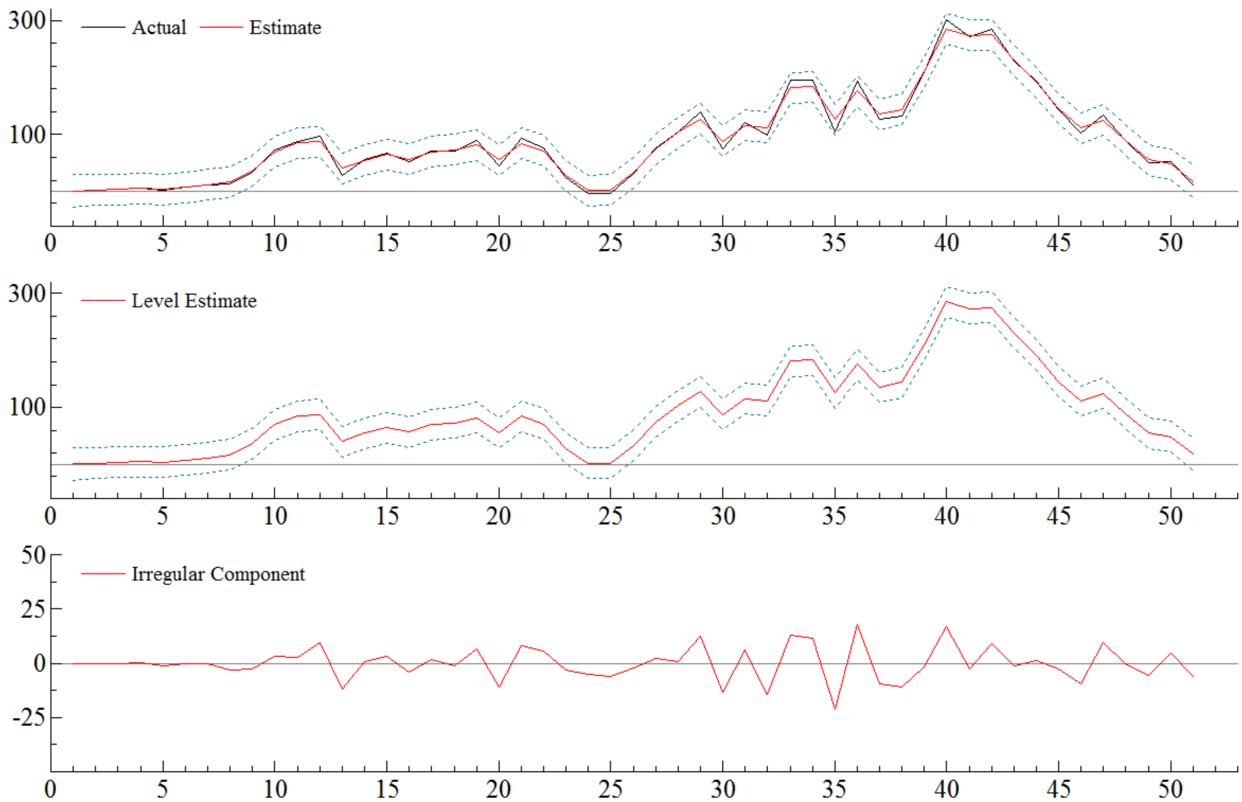

Figure 15: Fit of the Local Level Model.

To help choose between the three models suggested as candidates by our framework, Table 1 summarizes the fit statistics.

Model	Log-Likelihood	# of Parameters	AIC	BIC	p.e.v.
Local Linear Trend	-185.543	5	381.09	390.75	1758.65
Stoch. Level/Det. Trend	-185.553	4	379.11	386.83	1717.24
Local Level	-186.699	3	379.40	385.19	1715.98

Table 1: Fit Statistics for Estimated State-Space Models.

Both information criteria suggest that the unrestricted starting model (i.e. the local linear trend model) is the worst of the three. The extent to which it better fits the data is not large enough to justify the use of additional parameters. The choice between the local level model and the stochastic level/deterministic trend model is less clear – while the former is favoured by the BIC (and the prediction error variance), the AIC suggests that we select the latter. As the absolute value of the deterministic trend in the random-walk-plus-drift model is very close to zero and as the AIC has the tendency to select overparameterised models (Sneek, 1984), we proceed with the local level specification as our final model.

6.2 Residual Diagnostics

With linear Gaussian state-space models, the assumptions that need to be tested are (Commandeur and Koopman, 2007):

- **Independence:**

  We noted from the plot of the irregular component of the local level model that there was no apparent systematic pattern in the residuals, i.e. they seem to approximate White Noise. The ACF of the standardized residuals is plotted in Figure 16. It can be seen that at none of the plotted lags is the autocorrelation significantly different from zero.[24] The Ljung-Box statistic for lag 7 returns a value of $5.06 < \chi^2_{(6,0.05)} = 12.59$, suggesting that the assumption of independence is satisfied.

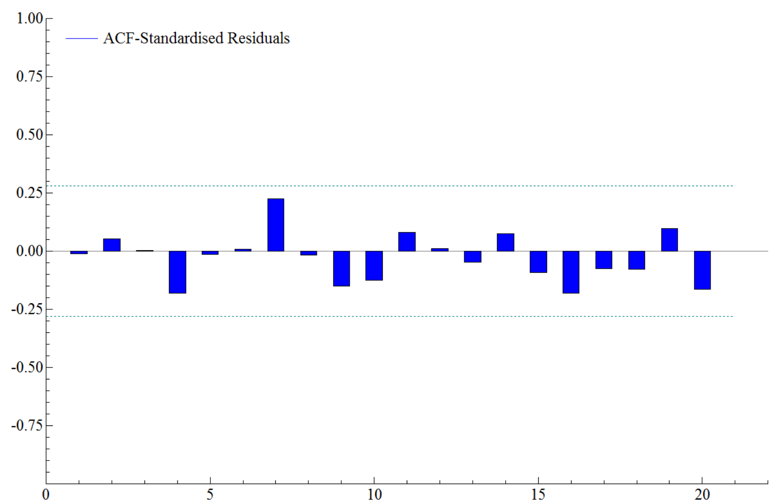

Figure 16: ACF of the Standardized Residuals (Local Level Model).

- **Homoscedasticity:**

  The following non-parametric test statistic for homoscedasiticity was used:

$$H(h) = \frac{\sum_{t=n-h+1}^{n} e_t^2}{\sum_{t=d+1}^{d+h} e_t^2}$$

  where d is the number of diffuse initial elements, n is the total number of observations, e_t^2 is the t^{th} residual and h is the closest integer to $(n-d)/3$. For our model, the test returns a value

[24] The 95 % confidence interval is given by $\pm 2/\sqrt{n} = \pm 0.28$.

of $H(16) = 5.1952$, which is larger than the 95 % critical value of $F(16,16;0.025) = 2.76$. This suggests that the variance of the residuals is increasing with time – heteroscedasticity is present. The estimated standard errors are inefficient and should be used with caution. It would be advisable to use robust White standard errors, or to use Weighted Least Squares (WLS) for estimation (Harvey, 1989, Chapter 3.4.2) – however, this is not currently implemented in STAMP.

- **Normality:**

  The histogram and quantile-quantile plot in Figure 17 return a good approximation to normality. The Bowman-Shenton test statistic of 0.63 is considerably lower than the relevant 95 % critical value of $\chi^2_{(2,0.05)} = 5.99$.

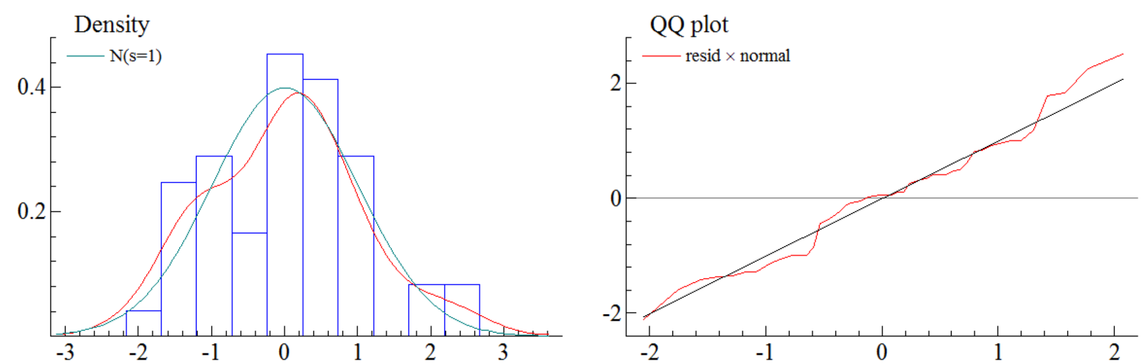

Figure 17: Histogram and QQ-Plot of the Standardized Residuals (Local Level Model).

	Statistic	Value	Critical Value	Assumption Satisfied
Independence	r(1)	-0.01	±0.28	+
	r(7)	0.23	±0.28	+
	Q(7)	5.06	12.59	+
Homoscedasticity	H(16)	5.20	2.76	-
Normality	N	0.63	5.99	+

Table 2: Residual Diagnostics (Local Level Model).

The residual diagnostics are summarized in Table 2, and represent a considerable improvement over the diagnostics obtained for the differential equation-based models. Overall, there are no major issues. The results of the non-parametric test for heteroscedasiticity is only a minor reason for concern – the Maximum Likelihood estimates are unbiased even under heteroscedasticity. It would nevertheless be good to use robust standard errors in estimation to obtain efficient estimates.

6.3 Bayesian Estimation

As an alternative to the frequentist approach so far, state-space models can be estimated using Bayesian methods. Petris and Petrone (2011) suggest that estimation with an MCMC algorithm might be more suitable for state-space models, compared to Maximum Likelihood estimation which does not allow for uncertainty about the true values of the model parameters.

We use a Gibbs sampler for the MCMC estimation of the model hyperparameters and use an inverse-gamma distribution as prior distribution for both hyperparameters. The inverse-gamma distribution with low shape and rate parameters is a commonly used non-informative prior distribution for variance parameters (Gelman, 2006).[25] Hence, the prior distributions are specified as follows: $(\sigma_{\varepsilon}^2){-1} \sim \mathit{Gamma}(\alpha_{\varepsilon},\, \beta_{\varepsilon})$ and $(\sigma_{\xi}^2){-1} \sim \mathit{Gamma}(\alpha_{\xi},\, \beta_{\xi})$, where α and β are the shape and the rate parameters of the Gamma distribution, respectively. Consistent with past research in the field (e.g. Petris et al., 2009), we set all shape and rate parameters to 0.01.

The algorithm was specified to draw 100,000 samples from the joint posterior distribution and save every tenth sample (i.e. thinning = 9) in order to avoid autocorrelation problems. Also – as common in MCMC analysis – we discarded the initial 1,000 samples as burn-in in order to reduce the influence of the starting value.

Figure 18 shows the results of the MCMC iterations. The plot of the running ergodic sample means suggests that convergence (towards a stationary distribution) was reached for the estimates of both parameters (the dashed red lines represent their posterior expectations). Despite the thinning procedure, the autocorrelation is remains rather high and decays slowly (particularly for the observation variance σ_{ε}^2), which is not a major problem given the large number of samples generated in the simulation.[26] It is striking that the estimates are quite different from the Maximum Likelihood estimates (even though the MLEs still lie within the 95 % posterior probability intervals). The Bayesian estimates are 95.22 (with a 95 % posterior probability interval of 0.083 to 580.62) for $\hat{\sigma}_{\varepsilon}^2$, and 1677.80 (with a 95 % posterior probability interval of 832.22 to 2648.28) for $\hat{\sigma}_{\xi}^2$.

[25] For a skeptical discussion of the use of inverse-gamma distributions as non-informative priors, see Gustafson et al. (2006).

[26] Attempts to increase the thinning factor did not have a strong effect on the estimates. Given the critical views on the practice of thinning – and, in particular, over-thinning – in the Bayesian literature (e.g. Jackman, 2009; Christensen et al., 2011; Link and Eaton, 2012), with the argument that autocorrelation is not fatal in itself, as long as a long run of the sampler is used, and that thinning of chains is not usually appropriate when the goal is precision of estimates from an MCMC sample, we refrain from increasing the thinning factor further.

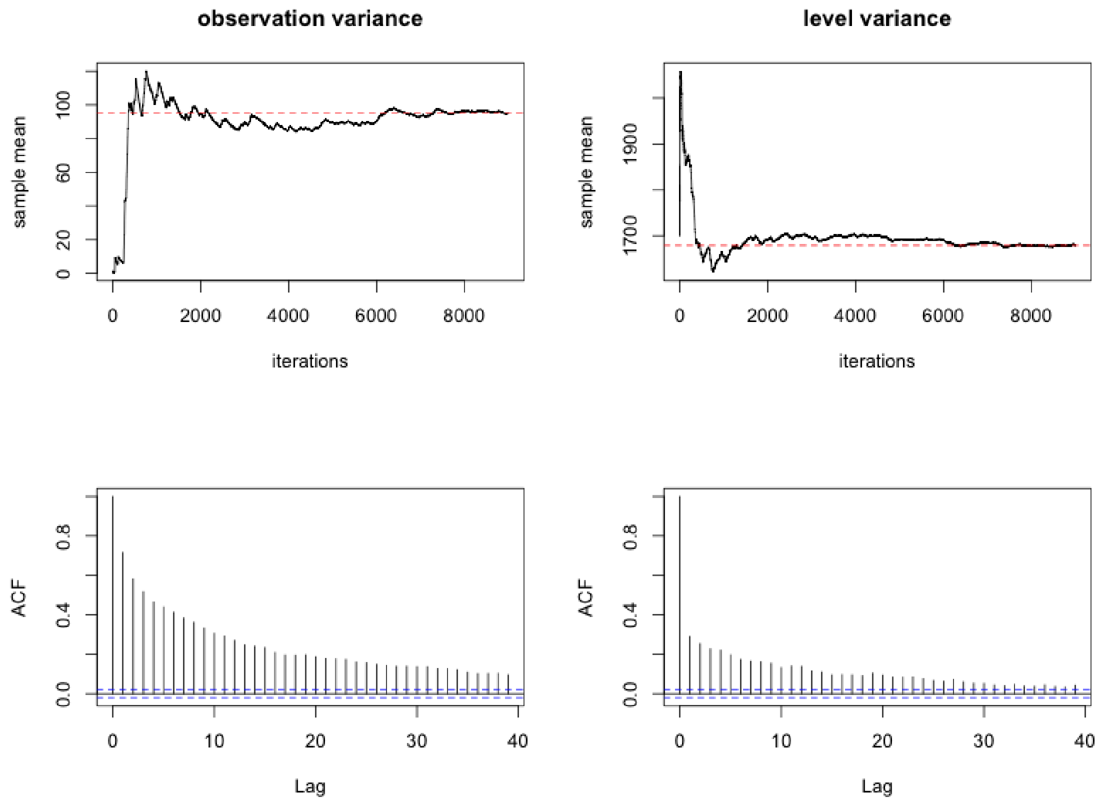

Figure 18: (above) Running sample means from MCMC iterations; (below) ACF for the estimates of the hyperparameters.

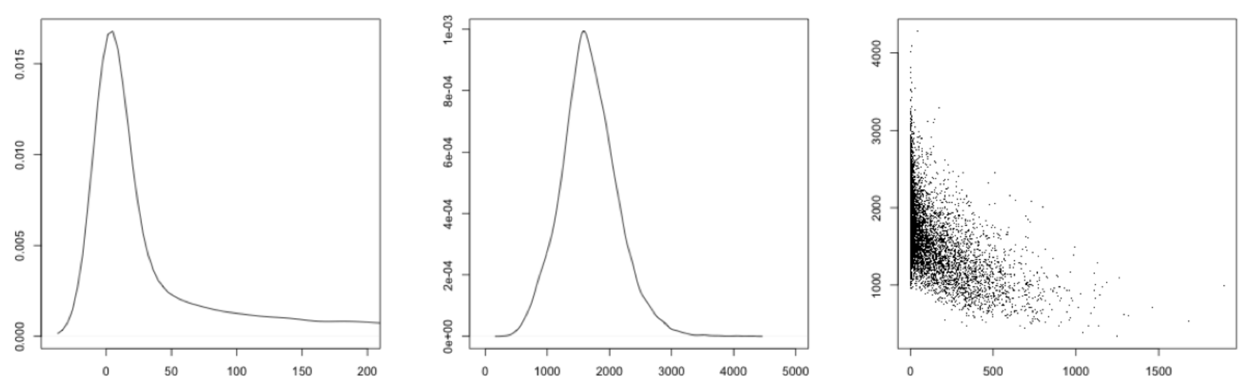

Figure 19: (a) Estimated Posterior Density of σ_ε^2; (b) Estimated Posterior Density of σ_ξ^2; (c) Samples from the Joint Posterior Density.

The most plausible explanation for the big differences between Bayesian estimates and Maximum Likelihood estimates is that the asymptotic Normal distribution of the Maximum Likelihood estimate does not approximate its actual sampling distribution very well (Petris and Petrone, 2011).[27] This coincides with the evidence in Figure 19 – the MCMC estimates of the posterior den-

[27]For a general overview of the regularity conditions that ensure asymptotic normality and consistency of the MLE,

sities of the state disturbance variances (in particular the one for σ_ε^2) show clear divergence from normality. The third panel plots the MCMC samples from their joint posterior density – the high correlation that is evident reflects a rather slow mixing of the Gibbs sampler.

7 Evaluation of the Models

7.1 Goodness of Fit

To assess the goodness of fit of the different models we use measures that have been extensively used in recent empirical diffusion studies (e.g. Michalakelis et al., 2008; Jukic, 2013) – the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), the Mean Percentage Error (MPE) and the Mean Absolute Percentage Error (MAPE).[28] The first two measures are scale-dependent, and the latter two are scale-independent (Hyndman and Koehler, 2006). As we only work with one data series in this report, scale dependency does is not an issue – however, for the sake of generalizability of the methods and tests we include scale-independent fit statistics as well.

The fit statistics of the estimated models are presented in Table 3.[29] Among extant diffusion models, the most striking feature is that the random walk (i.e. ARIMA) model appears to fit the data considerably better than the models based on differential equations. Within the Bass framework, the NLS method produces a slightly better fit than OLS estimation, although the OLS has a lower mean absolute error. The percentage-based error measures show that the fitted values are quite different from the actual values. The same is true for the logistic model which appears comparable with the Bass models in terms of fit. This is not surprising, as differential equation-based models seek to capture the long-term behavior of the diffusion curve, while the ARIMA model is better suited to capture shocks and temporary breaks.

It can be seen that the state-space models fit the data considerably better than the ARIMA model. This is not unexpected, as the state-space framework accommodates stationary and non-stationary features (Cowpertwait and Metcalfe, 2009). It is clear that state-space models offer a suitable way to describe diffusion paths of innovations. We proceed to examine the usefulness of the framework for forecasting, which is a key purpose of empirical diffusion modelling.

see Greene (2012); for a discussion in the context of structural time series models, see Harvey (1989), Chapter 4.5.1.

[28] $MAE = \frac{1}{n}\sum_{i=1}^{n} \mid \hat{x}_i - x_i \mid$, $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}$, $MPE = \frac{100}{n}\sum_{i=1}^{n} \frac{\hat{x}_i - x_i}{x_i}$, $MAPE = \frac{100}{n}\sum_{i=1}^{n} \mid \frac{\hat{x}_i - x_i}{x_i} \mid$.

[29] We present the fit statistics for the incremental data series (i.e. the number of new tractors), since this was the series that was modelled by the Bass and ARIMA models. The fitted values of the logistic model were transformed to first difference, in order to compare the model values with the actual incremental series.

Model	MAE	RMSE	MPE	MAPE
Bass (OLS)	37.99	50.62	132.40	207.60
Bass (NLS)	38.18	47.00	79.76	125.08
Logistic	41.36	53.12	46.26	212.78
ARIMA	32.06	41.90	7.64	68.08
State-Space (Local Level)	4.55	5.92	2.35	10.87
State-Space (Local Linear Trend)	4.11	5.43	1.41	9.79

Table 3: Fit Statistics for the Estimated Models.

7.2 Forecast Quality

In evaluating forecast quality we take into account the differences between the different model frameworks, i.e. the long-term focus of the differential equation-based methods and the short-term focus of the time-series based methods. We evaluate overall forecast quality using both short-term and and medium- to long-term forecast accuracy. To assess the short-term accuracy, we examine one-step ahead pseudo out-of-sample forecasts at different points on the diffusion curve (models re-estimated with truncated data and forecast for one period ahead). To assess medium- to long-term forecast accuracy we forecast for last 5 years, using models re-estimated using the first 46 data points. Not all of the modelling approaches provide standard errors[30], and so we use point estimates for forecast evaluation.

- ***Short-Term Forecasting:*** We re-estimated the models using data for the first 15, 20, 25, 30, 37, 45, and 50 years, respectively.[31] The logistic growth model was not able to produce point estimates for the entirety of these windows and so is left out in the evaluation of short-term forecast accuracy. As a differential-equation based model (which emphasises the long-term behavior of the diffusion curve), its results can be expected to be similar to those of the Bass models – this holds true for those windows for which the logistic model converged.

[30]See Section 3.2 for limitations of the Bass OLS estimation.

[31]The non-linear least squares estimation algorithm showed convergence problems for periods below 15 years, as well as for the estimation at 35 and 40 years, respectively.

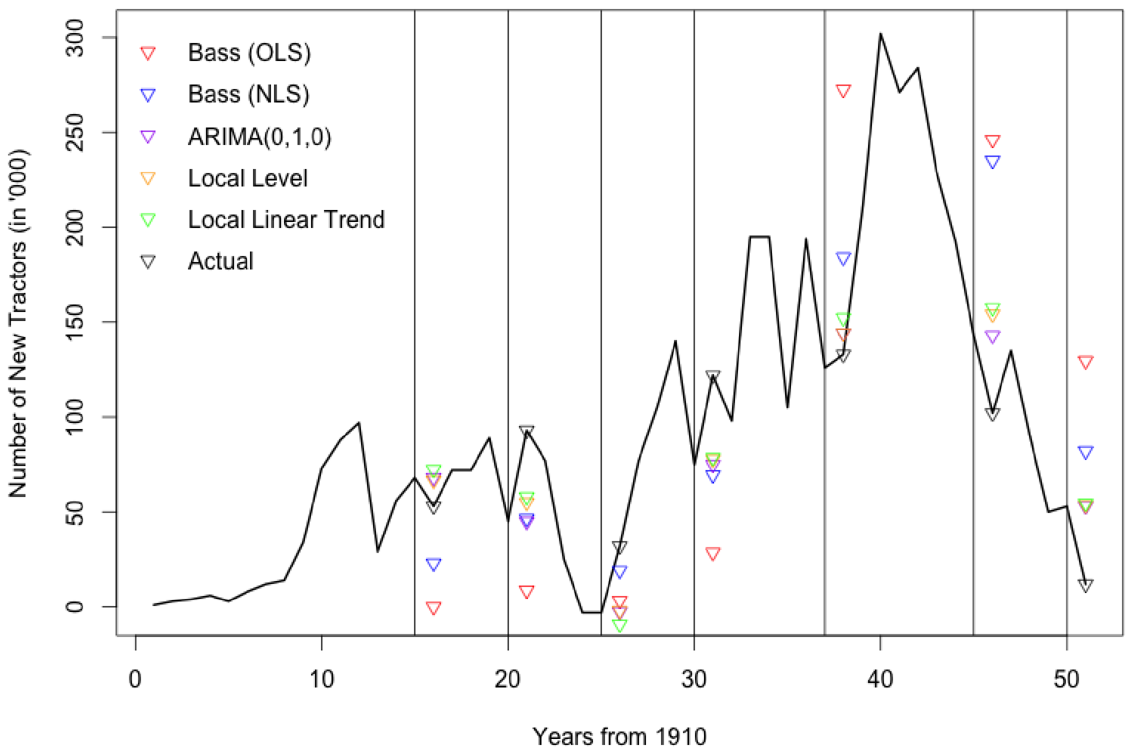

Figure 20: Selected One-Step Ahead Forecasts for the Estimated Models (Pseudo Out-of-Sample Forecasts).

The one-step ahead forecasts are shown in Figure 20 – the vertical lines indicate the time points at which the the data were truncated for re-estimation and one-step-ahead forecasts. Both Bass models underestimate the level for early periods and overestimate it for later periods (the NLS estimates of the Bass model produces considerably better forecasts than the OLS estimates). The superiority of the time-series methods for short-term forecasts is evident.

Table 4 presents one-step ahead forecasts produced by the models for given windows, as well as their deviations from actual values. The superiority of the time-series methods over the differential-equation based methods is obvious. It is also clear that adding a stochastic slope component to the state-space model does not add value for modelling this data series.

Period	Actual	Bass OLS	Bass NLS
15y	53	0.01 (-52.99)	23.05 (-29.95)
20y	93	8.76 (-84.24)	46.56 (-46.44)
25y	32	3.32 (-28.68)	19.20 (-12.80)
30y	122	28.71 (-93.29)	69.44 (-52.56)
37y	133	272.50 (+139.50)	184.21 (+51.21)
45y	102	246.01 (+144.01)	235.14 (+133.14)
50y	12	129.69 (+117.69)	82.12 (+70.12)

Period	Actual	ARIMA	Local Level	Local Linear Trend
15y	53	68 (+15)	66.68 (+13.68)	72.34 (+19.34)
20y	93	45 (-48)	55.05 (-37.95)	58.02 (-34.98)
25y	32	-2.56 (-34.56)	-1.98 (-33.98)	-9.08 (-41.08)
30y	122	75 (-47)	77.72 (-44.28)	78.62 (-43.38)
37y	133	144.15 (+11.15)	144.15 (+11.15)	152.22 (+19.22)
45y	102	143 (+41)	154.10 (+52.10)	157.52 (+55.52)
50y	12	53 (+41)	53.56 (+41.56)	54.57 (+42.57)

Table 4: One-Step Ahead Forecasts for the Estimated Models (Deviation from Actual Values in Brackets).

The importance of short-term forecasting cannot be overstated. Accurate short-term forecast help producers of innovations to plan their marketing and promotion efforts better, a sine qua non for competitive survival.

- ***Medium- to Long-Term Forecast:*** We re-estimated the models using the first 46 years of the observed series as input, and generated multi-step ahead forecasts for the last 5 years of the observation period. In fast moving markets for innovations, 5 years is a long time. The results do not change qualitatively when the forecast window is 10 years.

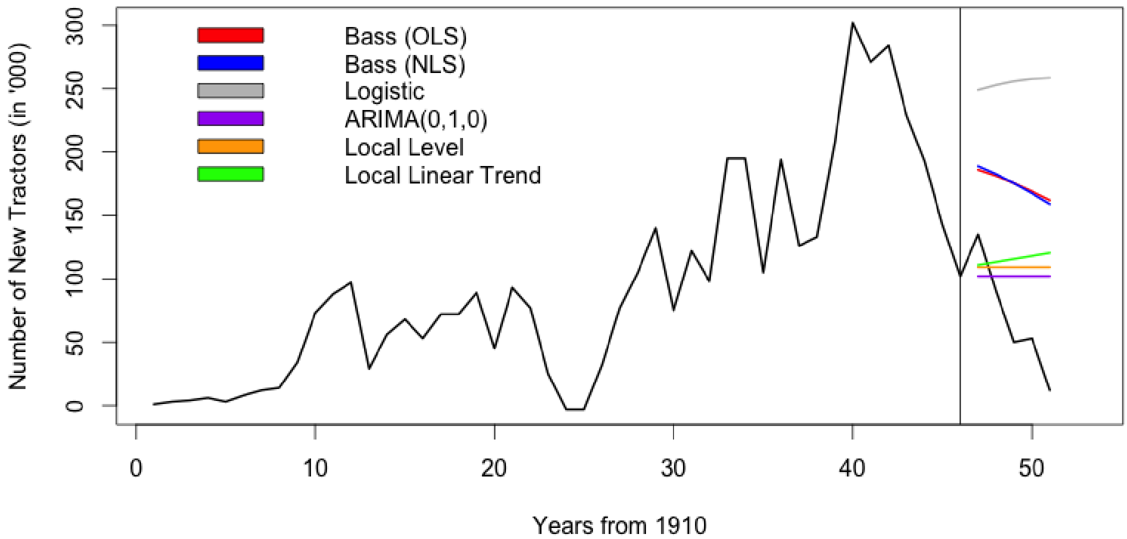

Figure 21: 5-year Ahead Forecasts for 1956 - 1960 for the Estimated Models (Pseudo Out-of-Sample Forecasts).

The results in Figure 21 have several notable features: First, the differential equations-based models (with the exception of the logistic growth model) tend to capture the long-term curve behavior – the decrease in the level in the later phase of the observation period – better, while the time-series methods are sensitive to the last observed value(s) of the series. Secondly, time-series methods appear to produce better long-term forecasts than the differential equations-based models. Thus even though the Bass model estimated using OLS and NLS, and the logistic growth model, are focussed on the long-term diffusion path, the time-series methods with their short-term focus produce better long-term forecasts due to their superior goodness of fit. Finally, the variation in the forecasts from the time-series models is low – all of them produce good approximations of the actual series – while the differential-equation based models produce very different forecasts (the logistic growth model predicts much higher levels than the Bass model). These visually gathered impressions are confirmed by the accuracy statistics in Table 5.

Model	MAE	RMSE	MPE	MAPE
Bass (OLS)	106.73	111.90	-371.73	371.73
Bass (NLS)	106.65	111.25	-366.75	366.75
Logistic	186.81	192.16	-623.69	623.69
Average DE-Based Model	133.40	138.44	-454.06	454.06
ARIMA	47.20	53.74	-187.07	196.85
State-Space (Local Level)	51.48	58.51	-207.13	214.79
State-Space (Local Linear Trend)	57.28	65.42	-232.99	240.15
Average Time-Series Model	51.99	56.14	-209.06	217.26

Table 5: Accuracy Statistics for the Long-Term Forecasts.

Their dominance over differential equation-based models in both short-term and long-term forecasts make a strong case for time-series based methods. While the interpretability of the Bass parameters (and parameters from the logistic growth model to a lesser extent) are appealing, and these models provide estimates of the long-term adoption paths, operational (short-term) decisions will be better served by the more accurate forecasts from time-series models. It is a cause for concern that the estimated Bass parameters are sensitive to the number of observations used as input (Van den Bulte and Lilien, 1997) – an OLS Bass model estimated with the first 45 years as input gives the upper market boundary as 10656.75; while with the first 50 years as input, the estimated upper market boundary nearly halves to 5500.66. The dependence of estimates upon the point on the diffusion path where the data are truncated casts doubts on the usefulness of the estimated Bass parameters. However, it might be possible to exploit the data dependence of Bass parameters to analyze the timing of the takeoff, an aspect of diffusion that is of much interest (e.g. Golder and Tellis, 1997; Chandrasekaran and Tellis, 2008; Peres et al., 2010). By re-estimating the Bass model at multiple points on the diffusion curve with truncated data, the sudden and sharp transition from the slow initial uptake to the fast (superexponential) growth could be revealed by changes in the estimates of the upper market boundary and/or the coefficient of imitation.

8 State-Space Models with Regressors and Intervention Variables

8.1 Background & Modelling Framework

In Section 6, we established that simple linear Gaussian state-space models that use information on the dependent variable alone can characterise diffusion processes better than the methods in standard use. Furthermore, the state-space approach can accommodate intervention variables and/or regressors. As discussed in Harvey (1989), and Commandeur and Koopman (2007), intervention variables can capture three dynamic developments in non-seasonal time series:[32]

- **Level Shift:** A sudden change of the level of the time series, where the level change is permanent after the intervention.

- **Slope Shift:** A sudden significant change of the slope of the time series, where the slope change is permanent after the intervention.

- **Pulse** (or 'transitory effects', as denoted by Harvey, 1989)**:** A significant change of the level of the time series at the point of intervention, where the level immediately returns to its pre-intervention value at the following point in time.

Depending on whether the causes of potential interventions are known to the analyst, the intervention points can be chosen manually (e.g. points of policy introductions or changes) or automatically (i.e. data driven selection of intervention points allowing for the inherent dynamics of the time series).

Causal factors relevant to a diffusion process can be introduced into the analysis through regressor variables. Unlike ARIMA models that cannot accommodate explanatory variables, or the Bass framework that is not very flexible in accommodating regressors[33] and does not allow for intervention effects at all, these elements can be introduced into state-space models in straightforward ways.

[32] In the case of seasonal time series, a change in the seasonal pattern could also justify the use of an intervention variable (Harvey, 1989).

[33] A generalized Bass framework that uses decision variables has been proposed by Bass et al. (1994).

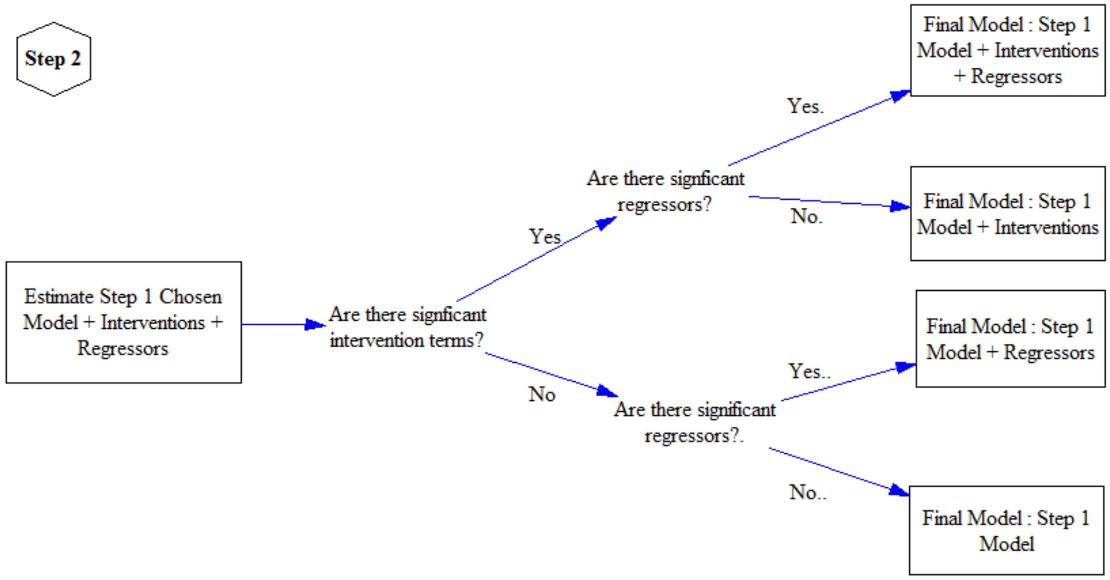

Figure 22: Proposed Approach for State-Space Model Selection in a Diffusion Context (Case With Interventions and/or Explanatory Variables).

The model selection framework that was proposed in Section 6 can be extended as follows (Figure 22):

1. Choose the final model from Step 1 (without regressors or interventions) as the starting model. If we had concluded that a deterministic level/slope was sufficient at that stage, then the introduction of additional determinants of the process should not change that conclusion. The newly introduced determinants may explain part of the dynamics of the diffusion process and thus decrease a part of the variation that was attributed to the state variables. Hence it is not necessary to repeat the entire Step 1.[34]

2. Estimate the final model from Step 1 including interventions (automatically or manually selected) and regressors of interest.

3. Examine significances of the interventions and the regressors and reduce the model accordingly to obtain the final model.

The inclusion of regressors and interventions in the state-space framework provides the decomposition of the dynamic effects that drive the diffusion process. Unless the model is intended

[34] The reason to fit a univariate structural time series model to the dependent variable (which can highlight the salient dynamic features in the data), which can in turn be used to inform a ‘richer’ model that includes explanatory variables is based on the discussion by Harvey (1989), Chapter 7.5.

only for forecasting, the final model from Step 2 should be seen as independent from the final model from Step 1 in the sense that Step 1 provides a model that identifies the inherent dynamics of the data (internal dynamics), while Step 2 provides a model that can accommodate a theoretical/causal/structural basis (internal & external dynamics).

8.2 Empirical Analysis

For the empirical analysis of the tractor data, we use regressors used in Manuelli and Seshadri (2014) – namely:

- **Tractor Price:** The (quality-adjusted) real tractor price is expected to be negatively related to the number of new tractors in use.
- **Horse Price:** The real horse price is expected to be positively related to the number of new tractors in use, since horses (and mules) can be seen as a substitute 'technology' to tractors.
- **Wage Rate:** The real wage rate is expected to be positively related to the number of new tractors in use.
- **Costs of Tractor Usage:** The costs of using tractors on farms is expected to be negatively related to the number of new tractors in use.

Allowing for automatic selection of interventions, a pulse effect in period 35 (i.e. 1944) and a level shift in period 40 (i.e. 1949 and following) were identified.

The local level model (the final model from Section 6) was estimated with the above regressors and intervention variables (log-likelihood at convergence = -193.64). The Maximum Likelihood estimates of the hyperparameters were $\hat{\sigma}_{\varepsilon}^2 = 299.37$ (variance of the irregular component) and $\hat{\sigma}_{\xi}^2 = 1209.25$ (variance of the level disturbance), respectively. The Maximum Likelihood estimate of the initial value of the level was $\hat{\mu}_1 = -120.30$ (in this model, the stochastic level 'regulates' the fixed regression effect).

Table 6 shows the regression output for the local level model with interventions and regressors. The directional effects of the explanatory variables are as expected – higher wage and higher horse prices lead to an increase in the number of new tractors used, while higher tractor price lead to a decrease in the number of new tractors used (ceteris paribus). The effect of the costs of tractor usage is negative but insignificant.

Regressor	Estimate	Standard Error	p-value
Pulse (Period 35)	-110.536	33.696	0.002
Level Shift (Period 40)	108.265	42.079	0.014
Tractor Price	-0.266	0.151	0.092
Horse Price	5.120	2.733	0.064
Wage Rate	0.505	0.194	0.013
Cost of Tractor Usage	-0.266	0.626	0.672

Table 6: Regression Effects of the Local Level Model with Interventions and Regressors.

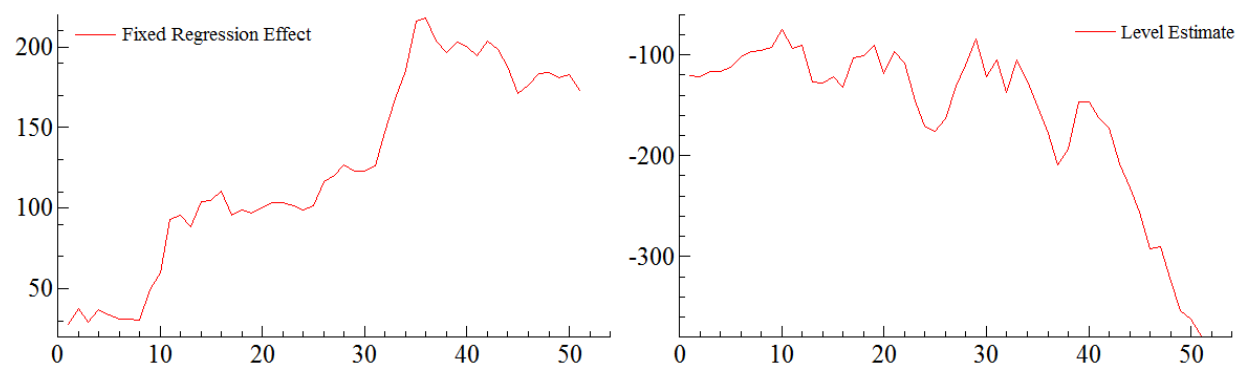

Figure 23: (a) Fixed Regression Effect of the Local Level Model with Interventions and Regressors; (b) Level Estimates of the Local Level Model with Interventions and Regressors.

Figure 23 plots the fixed regression effect for the entire observation period, as well as the stochastic level component of the model. In its growth phase the diffusion process seems to be well explained by the regressors (and interventions), but the regression effect does not fit the decline in the last few years at all. The decline phase is almost entirely captured by the stochastic level component. The declining growth rate in the last few years of the observation period can most reasonably be ascribed to saturation, which would be an inherent feature of a diffusion process dynamics rather than a feature that is driven by external variables. Overall, the model including regressors and interventions can explain the variation in the series rather well: an R^2 of 77.33 % (as opposed to 72.10 % for the local level model without regressors and interventions and 55.73 % for the Bass OLS model, respectively).

9 The Saddle Effect

9.1 Background

The stylized S-shaped cumulative adoption will generally mask several sub-processes in real world innovation diffusions. Two common phenomena that lead to deviations from the classical S-shape are the saddle effect, and technological substitution (Figure 24).

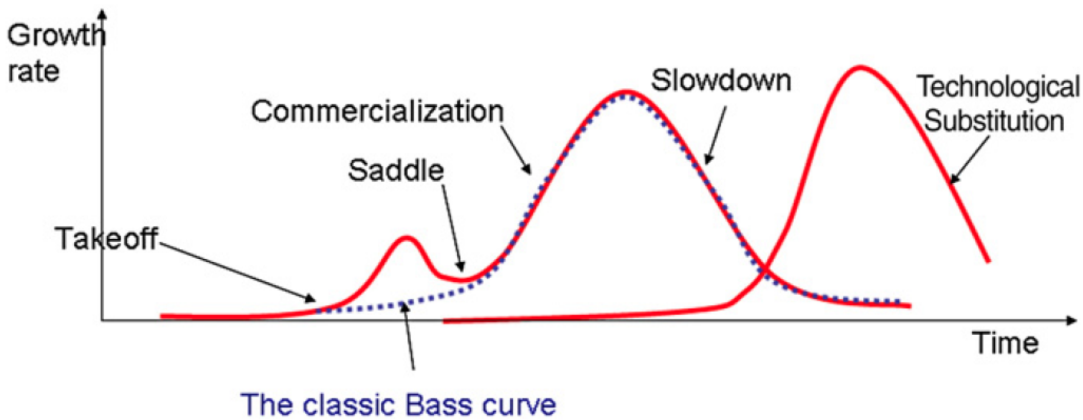

Figure 24: Sub-Processes of a Product Life Cycle (Peres et al., 2010).

The saddle effect has been of great interest in recent diffusion research.[35] In the marketing context the saddle has been characterised as a pattern in which an initial peak predates a trough of sufficient depth and duration which is followed by sales eventually exceeding the initial peak (Goldenberg et al. 2002). Saddle effects have been found in the diffusion patterns of over 50 % of product innovations (Goldenberg et al., 2002; Golder and Tellis, 2004).

One set of researchers, most notably Goldenberg et al. (2002), Muller and Yogev (2006), and Van den Bulte and Joshi (2007), base their explanations for the saddle upon the heterogeneity among potential adopters, in particular between early adopters and early majority – categories first described by Moore (1991). A saddle can arise from a communication gap between these adopter groups – then early adopters are not reliable drivers of the early majority's decision to adopt. Another explanation is based information cascades,[36] when many potential adopters base their adoption decisions on the behavior of few previous adopters rather than their individual evaluations of utility (Golder and Tellis, 2004). This can lead to sudden and sharp takeoff (positive cascade), but also make the market prone to a sudden decline in growth (negative cascade). Chandrasekaran and Tellis (2011)

[35] For an overview of studies that describe and/or analyse saddle effects, see Shankar and Carpenter (2012).
[36] Bikhchandani et al. (1992).

find empirical support for saddle effects in the diffusion of technological innovations caused by chasms in the group of potential adopters, by technological cycles, as well as by business cycles.

An unresolved issue in the research on the saddle effect is its identification. The characterisation in Goldenberg et al. (2002) – 'sufficient depth and duration' – is intuitive, but arbitrary. It is unclear how deep and long a saddle must really be to satisfy this condition. Goldenberg et al. (2002) classify a sub-process as a saddle if the depth of the decline is at least 20 % of the previous peak for a duration of at least two years. Similar definitions were used in Golder and Tellis (2004), and Stremersch and Tellis (2004). Goldenberg et al. (2002), and Chandrasekan and Tellis (2011), find an average duration of 8 years for saddles with a sales decline of 15 - 32 %. A more systematic approach to identification and inference will be of value.

9.2 Testing for a Saddle (State-Space Framework)

The state-space framework lends itself to a well-founded strategy for the identification of saddles in diffusion. As a state-space model offers a structural decomposition of a time series, examination of the separate components can be helpful in identifying the saddle. To avoid the misclassifying random fluctuations as saddles, it is reasonable to hold that the decline in growth must last a long enough period. At issue is the depth of the decline – a decline by 20 % in a process that is fluctuating a lot is different from such a decline in a relatively stable process. It is appropriate to take into account the dispersiveness in the process when considering the depth required to identify a saddle. In the state-space framework the confidence intervals around the point estimates of the stochastic components make this possible.

We define two different kinds of saddle effects:

- **Saddle in the narrow sense:** A significant decline in the cumulative adoption level for a period of at least 2 years that falls before the global inflection point.

- **Saddle in the wider sense:** A significant decline in the incremental adoption level for a period of at least 2 years that falls before the global inflection point.

The interpretation of the saddle in the narrow sense is straightforward – a decline in the cumulative adoption level indicates that the number of new adopters is more than offset by the number of previous adopters who abandon the innovation. The saddle in the wider sense can be explained as a deceleration of the adoption process – a decline in the incremental adoption level (i.e. the number of new adopters in a given period) indicates a lower adoption rate. It should be clear that by these definitions a saddle in the narrow sense is always also a saddle in the wider sense, but not the converse.

The reason for this differentiation is based on both technical and behavioral/structural considerations – first, a saddle in the narrow sense can easily be tested for by using the cumulative adoption level series, while a saddle in the wider sense can be tested for by using the incremental adoption level series. Secondly (and more importantly), it seems intuitive that the generating mechanisms for the two kinds of saddles are very different – while a deceleration of the adoption process (or even a complete stabilization of the process for a limited time period) can be caused by the heterogeneity in the population of potential adopters (and in particular by the chasm between early adopters and early majority), this phenomenon cannot explain a decline in the cumulative adoption level (i.e. a saddle in the narrow sense). The second generating mechanism discussed extensively in the literature – information cascades – could well explain a decline in the cumulative adoption level: if many adopters have based their decision to adopt on the behavior of few previous adopters rather than an individual evaluation of utility from the alternatives, it is possible that they re-evaluate alternatives at a later point and reverse their decisions (negative cascade effect). Quite naturally, reasons for an intermediate decline in the cumulative adoption level could also be overall economic or social conditions, or other events like the market entry of a rival innovation that can serve as a substitute (harsher competition) or adverse publicity in relation to the innovation. However, depending on how well the given diffusion process is understood, all reasons can be explored in the main modelling process of the state-space framework through explanatory variables or interventions.

For the identification of the two forms of the saddle, we propose the following framework (Figures 25 & 26):

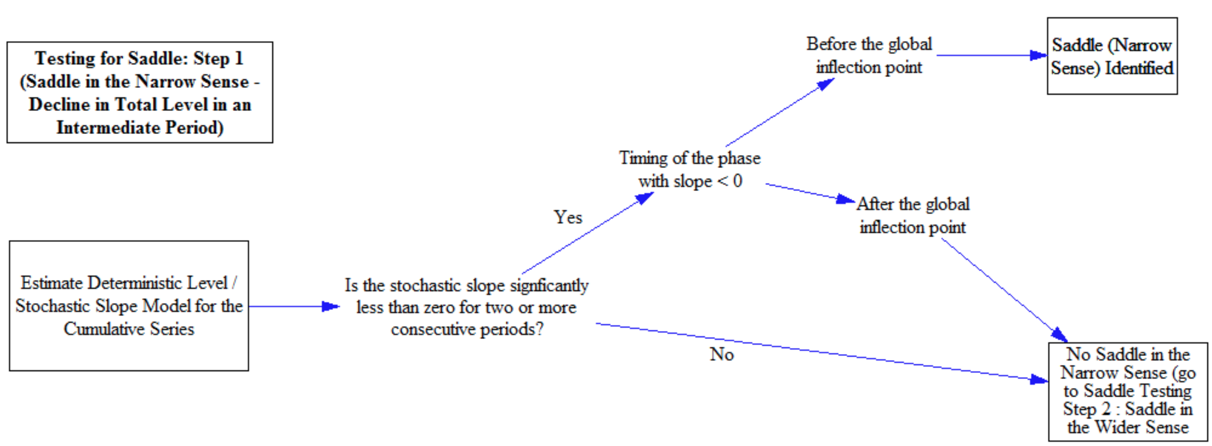

Figure 25: Proposed Framework for a Systematic Test for a Saddle in the Narrow Sense (Decline in Cumulative Adoption Level).

1. **Identification of the saddle in the narrow sense** (decline in total adoption level):

    - Estimate an auxiliary state-space model with only a deterministic level and a stochastic slope component for the cumulative series. For completed history diffusion processes, the estimate of the deterministic level will be close to zero, since the process starts with the market introduction of the innovation. It is not a problem if the process starts at a level different from zero, since we are interested in growth rather than the level (it should be clear from the discussion in the previous paragraph that a saddle in the narrow sense can be interpreted as negative growth).

    - Examine the stochastic slope component (which captures the entire growth effect in this specification of the state-space model) and its confidence bands.

        – If there is a period that shows a stochastic slope that is significantly lower than zero and located before the global inflection point, this constitutes a saddle in the narrow sense.

        – If there is no period that shows a stochastic slope that is significantly lower than zero or if such a period falls after the global inflection point, the diffusion process does not show signs of a saddle in the narrow sense. Proceed with Step 2.

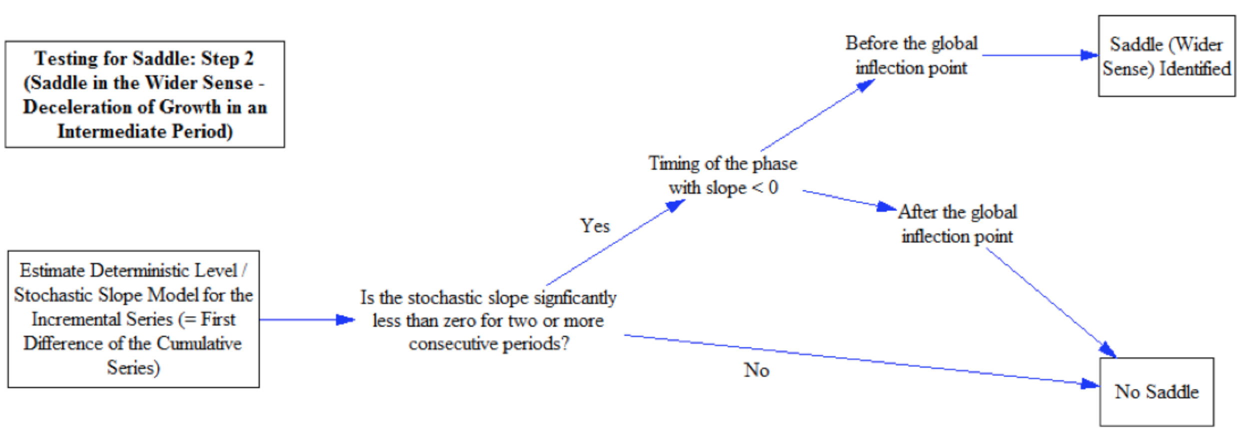

Figure 26: Proposed Framework for a Systematic Test for a Saddle in the Wider Sense (Deceleration of the Adoption Process).

2. **Identification of the saddle in the wider sense** (deceleration of the adoption process):

- Estimate an auxiliary state-space model with only a deterministic level and a stochastic slope component for the incremental series. Again, for completed history diffusion processes the estimate of the deterministic level will be close to zero, since the number of new adopters at the time of the innovation's market entry is zero. Again, as in the case of the saddle in the narrow sense, a non-zero starting level is not a problem, as long as a deterministic level component is used in our auxiliary model.

- Examine the stochastic slope component (which captures the entire growth effect in this specification of the state-space model) and its confidence bands.

  – If there is a period that shows a stochastic slope that is significantly lower than zero and falls before the global inflection point, this constitutes a saddle in the wider sense.

  – If there is no period that shows a stochastic slope that is significantly lower than zero or if such a period falls after the global inflection point, the diffusion process does not show signs of a saddle.

There are two challenges in this approach to saddle inference. First, the identification of a saddle depends heavily on the knowledge of the global inflection point (i.e. the global peak in the number of new adopters in a given period). This clearly makes the test more suitable for ex-post analysis of a completed diffusion process than for the analysis at intermediate points of an ongoing process. This is also a problem in the extant definition of the saddle effect. It makes intuitive sense that there can be no way of identifying a saddle without having a good idea about the global inflection point. This makes saddle analysis at intermediate periods in the diffusion process difficult. Secondly, the choice of the confidence level used must depend on the richness of the data. Usually time series data on innovation diffusion are not very long. It would be reasonable to report test results for a range of confidence levels, in order to get a good idea about the uncertainty of conclusions we can draw from a test. In the empirical analysis that follows, we tested for saddles at a 95 %, 50% and 10 % confidence level, respectively.

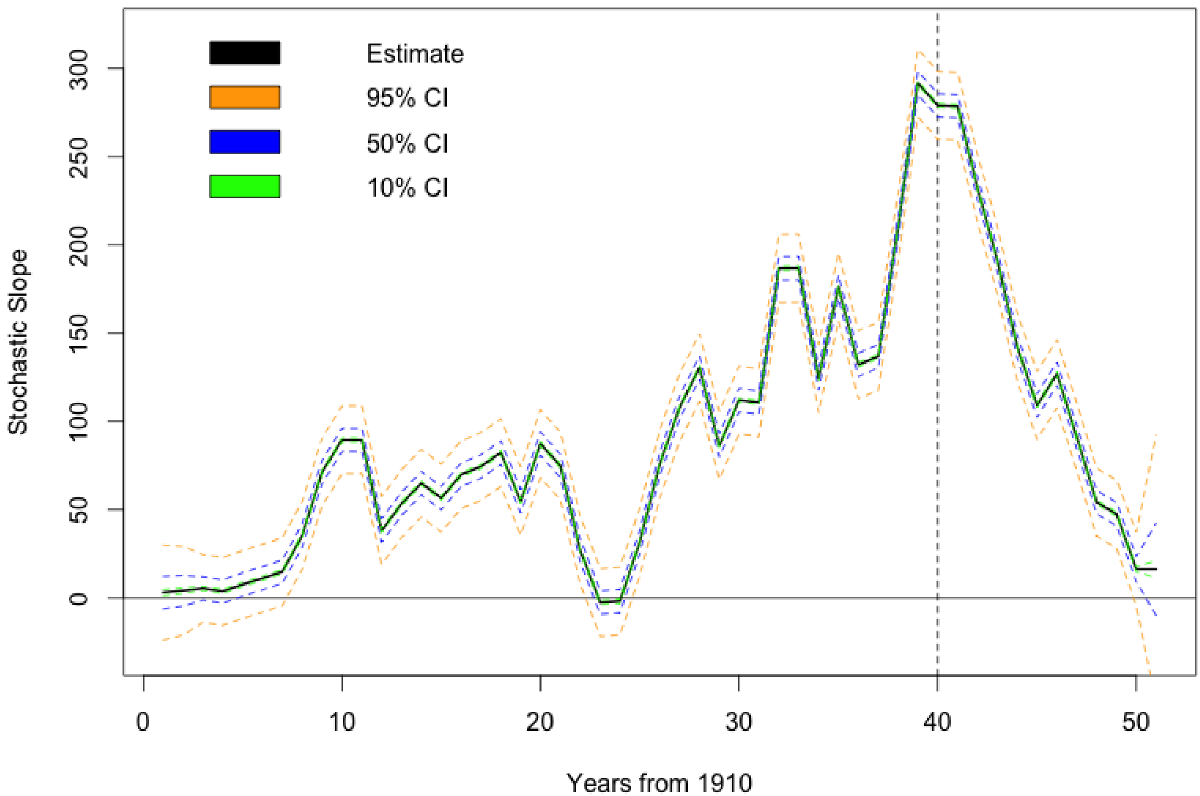

Figure 27: Stochastic Slope Estimates (Including Confidence Intervals) of the Auxiliary Model for the Cumulative Diffusion Series – Test for a Saddle in the Narrow Sense.

Figure 27 shows the stochastic slope estimates of the auxiliary deterministic level/stochastic slope model for the cumulative adoption time series. The estimate for the deterministic level is 0.949 – close to zero as expected. Since we examine a complete diffusion process that shows clear signs of saturation towards the end of the observation period, it is easy to identify the global inflection point. For this series it is at period 40 that the incremental adoption series reaches its peak. As can be seen in the figure, the only potential saddle in the narrow sense lies between periods 22 - 25. This is highlighted for visibility (Figure 28).

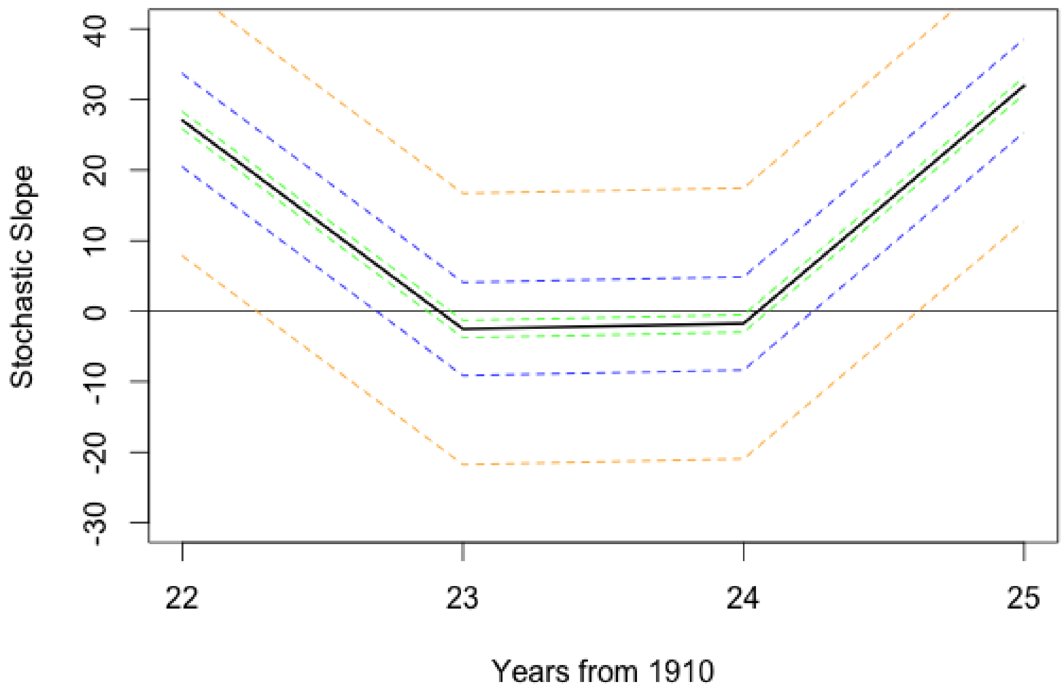

Figure 28: Zoomed-In Stochastic Slope Estimates (Including Confidence Intervals) of the Auxiliary Model for the Cumulative Diffusion Series for Period of Potential Saddle.

The graph shows that at a 10 % confidence level we can identify a saddle in the narrow sense (the estimates for two consecutive years – Periods 23 and 24 – are significantly lower than zero), while we find no evidence of a saddle in the narrow sense at the 50 % and 95 % confidence levels, respectively.

Figure 29 shows the stochastic slope estimates of the auxiliary deterministic level/stochastic slope model for the incremental adoption series. The estimate for the deterministic level is 0.964 – again very close to zero as expected. A first inspection of the stochastic slope estimates indicated the periods 11-15 and 18-24 as potential saddles in the wider sense, i.e. phases of significantly decelerating growth in the adoption process. The stochastic slope estimates including confidence bands are again zoomed in on to improve the visibility of the potential saddle effect (Figure 30).

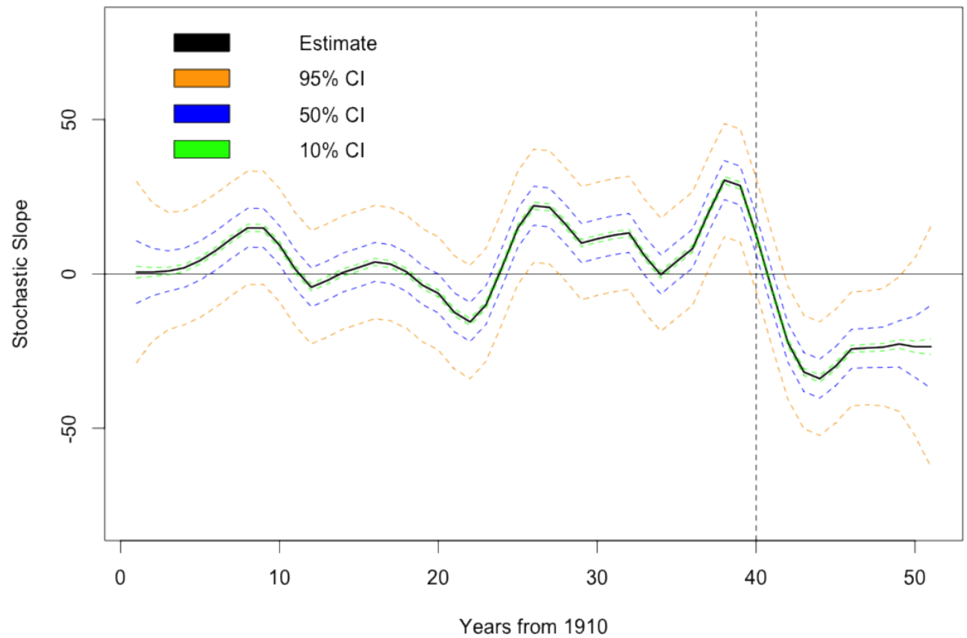

Figure 29: Stochastic Slope Estimates (Including Confidence Intervals) of the Auxiliary Model for the Incremental Diffusion Series – Test for a Saddle in the Wider Sense.

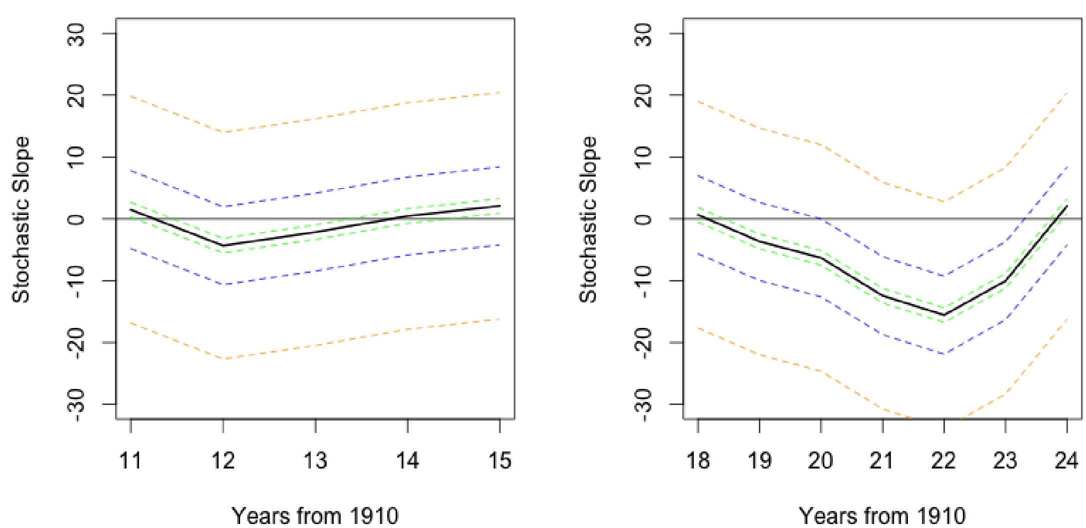

Figure 30: Zoomed-In Stochastic Slope Estimates (Including Confidence Intervals) of the Auxiliary Model for the Incremental Diffusion Series for Periods of Potential Saddles.

It can be seen that at 10 % confidence level, two saddles in the wider sense (i.e. two phases of significant growth deceleration) are identified – one at Periods 12-13 and one at Periods 19-23. At the 50 % confidence level, only one saddle in the wider sense is identified at periods 20-23. At the 95 % confidence level, no saddles in the wider sense are identified.

Confidence Level	Saddle in the Narrow Sense	Saddle in the Wider Sense
10 %	Periods 23-24	Periods 12-13 & 19-23
50 %	-	Periods 20-23
95 %	-	-

Table 7: Results of the Systematic Test for Saddles.

The results of the saddle tests are summarized in Table 7. The final choice of the required confidence level is left to the researcher and is dependent on the richness of the data. We consider a 95 % confidence level too strict given the limited length of most diffusion time series and consider a 10 % confidence level as too lenient for the identification of saddles – risking misclassification of random fluctuations as saddles. The 50 % confidence level seems a reasonable choice in the context of typical diffusion time series.

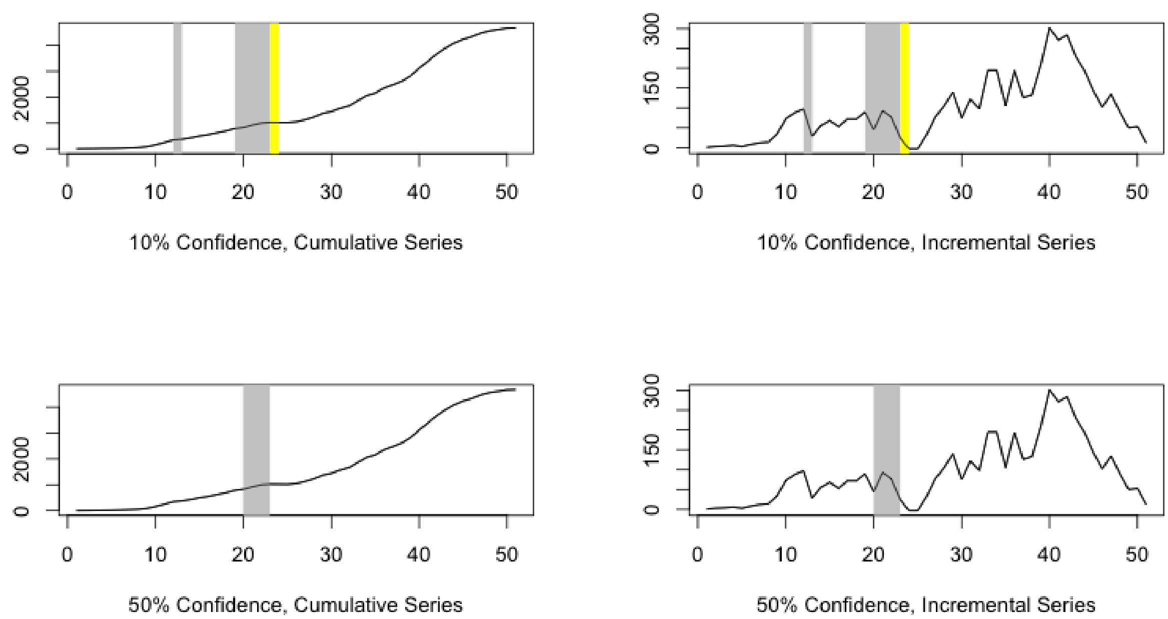

Figure 31: Identified Saddles at 10 % and 50 % Confidence Levels (Yellow = Saddle in a Narrow Sense; Grey = Saddle in a Wider Sense).

Figure 31 shows the saddle periods we identified at different confidence levels in the cumulative and the incremental series respectively (saddles in the narrow sense are shown in yellow, and saddles in the wider sense are shown in grey).[37] This visual inspection supports the use of the 50 % confidence level for saddle inference – the minimal decline in the total level in the early 1930s should probably not be seen as significant given its size; however, it seems intuitive that the adoption process has decelerated during this phase. This is exactly what our test suggests at the 50 % confidence level, while the 10 % confidence level suggests that there was a significant decline in the total level and the 95 % confidence level suggests there was not even a significant deceleration in the adoption process.

10 Discussion & Conclusion

We have shown that state-space models provide a powerful way to extract salient features of a diffusion curve, and thus explain the process. The ability to fit the process well is crucial not only for forecasting, but also from a causal / structural / explanatory point of view. Young (2009) classifies the underlying (internal) drivers of innovation diffusion (derived from the literature on marketing, sociology, and economics) as follows:[38]

- **Contagion:** In a diffusion process that is driven purely by contagion, the adoption decision is based on exposure to previous adopters.
- **Social Influence:** In a diffusion process that is driven purely by social influence, the adoption decision is based on social pressures and a conformity motive. People adopt when sufficient numbers of other people in the population have adopted – this is closely related to the bandwagon effect described in Abrahamson and Rosenkopf (1993), and the concept of information cascades, as described in Bikhchandani et al. (1992).
- **Social Learning:** In a diffusion process that is driven purely by social learning, the adoption decision is based on evidence (from empirical observations on outcomes among prior adopters) on the balance between benefits and costs of adoption. In this setting, heterogeneity of the population of potential adopters is relevant, as differences in the prior beliefs, the information-gathering process, as well as idiosyncratic costs, can influence the timing of adoption.

[37] The results for the 95 % confidence level are not plotted, since no saddles were identified.

[38] It should be noted, that Young (2009) focuses on adoption processes that are driven from within (internal dynamics), rather than processes that are driven by external shocks, such as price or quality changes (external dynamics).

Social learning, the mechanism based on utility maximization, is the most plausible mechanism from an economic point of view (Young, 2009). The other two mechanisms are based on exposure.[39] Young (2009) shows that the different mechanisms leave their on specific dynamic 'footprints' on the adoption path, resulting in diffusion trajectories that differ in shape and acceleration patterns. His main findings include:

1. Diffusion processes that are driven purely by contagion are characterised by an initial acceleration and a deceleration as saturation is approached, resulting in the stylised S-shaped curve. Such processes can never accelerate beyond the 50 % adoption level.

2. Diffusion processes that are driven purely by social influence can either decelerate or accelerate initially, but in the latter case they do so at a superexponential rate for some period of time.

3. Diffusion processes that are driven purely by social learning are characterised by a slow initial uptake, and can either decelerate or accelerate initially, and typically accelerate at a superexponential rate for some period of time. Due to the relevance of heterogeneity, intermediate phases of deceleration can be explained in this setting.

From our empirical analysis, the internal dynamics of the diffusion of tractors appear to be described well by a social learning mechanism, as the series show slow initial adoption, a 'take-off' characterised by a superexponential acceleration, as well as intermediate phases of significant growth deceleration according to the results of our saddle test.

The issues raised and the results presented in this paper set a broad agenda for future work: a major limitation is to do with data. Using larger data sets with wider ranges of diffusion experience, will enable the exploration of the validity of the different mechanisms: social learning, contagion, and social influence. It is reasonable to conjecture that social learning will be a common mechanism in the diffusion of product and technological innovations, and that ideas and practices conceivably diffuse in different ways, since the cost/benefit calculus is arguably more remote in these contexts. Larger data sets should facilitate the evaluation of the forecasting power of state-space models in the diffusion context – our results show that the local linear trend model forecasts an increase in the incremental adoption level over the last 5 years of the observation period despite clear signs of saturation at this point. This was due to very limited variation in the slope, leading the local linear trend model to reduce to a random-walk-with-drift model (stochastic level/deterministic slope model). It remains to be explored whether this is a common issue when forecasting diffusion processes with

[39] For extensive discussions of the impact of social learning on diffusion processes, see e.g. Kapur (1995), Chatterjee and Xu (2004), Munshi (2004), Young (2005, 2009), and Lopez-Pintado and Watts (2008).

state-space models. Among practitioners a main focus is on establishing the market potential very early in the diffusion process (Meade and Islam, 2006). An open question is to do with the ability of state-space models to forecast diffusion rates with limited data.

Another direction of future research is the analysis of the timing of 'take-offs' – a topic that of great interest among practitioners as well as academics (e.g. Golder and Tellis, 1997; Golder and Tellis, 2004; Chandrasekaran and Tellis, 2008). As with saddles, take-offs have tended to be defined in an arbitrary ways for empirical purposes. It should be possible to develop systematic inference procedures for the timing of the take-off, exploiting the dynamic components of state-space models in a similar fashion to the saddle inference framework presented in this paper. Auxiliary residuals should be useful in exploring potential breaks in the slope (slope shifts).

Finally, an useful direction for future research would be the incorporation of network features into the state-space framework as explanatory variables. When analyzing the diffusion of ideas or policies, the effect of network topology (and consequently features such as centrality and density) on diffusion rates is of great interest (e.g. Valente, 1995; Bala and Goyal, 1998; Morris, 2000; Young, 2003; Jackson and Yariv, 2007; Jackson and Rogers, 2007; Delre et al., 2010; Golub and Jackson, 2010; Jackson, 2010; Kuandykov and Sokolov, 2010; Montanari and Saberi, 2010).[40] Incorporating such features into state-space models as explanatory variables might be a way to exploit the merits of both aggregate, and disaggregate structural/behavioral approaches to diffusion modelling, thereby refining the purely empirical approach of the basic state-space framework.

Overall, we have established that state-space modelling is an appealing and powerful approach to study diffusion phenomena. Of course there will be no single diffusion model that is best for all purposes. The proposed time-series methods, in particular the flexible state-space framework, should not be seen as substitutes for the classical diffusion models (Bass, Logistic, and Gompertz). A useful strategy will be to use the models as complements. For example, dynamic evolution of the Bass parameters (obtained from rolling estimates of the Bass model) might hold useful insights about diffusion features such as take-offs and saddles. Evidence gained from such analysis can be combined results from state-space models to make conclusions about take-offs and saddle effects more robust.

[40] However, note that recent research by Kreindler and Young (2013), disputes the relevance of network topology in the diffusion process by constructing an agent-based diffusion model without constraints on network topology.

References

Abrahamson, E., & Rosenkopf, L. (1993). Institutional and Competitive Bandwagons: Using Mathematical Modeling as a Tool to Explore Innovation Diffusion. *Academy of Management Review, Vol. 18, No. 3,* pp. 487-517.

Abrahamson, E., & Rosenkopf, L. (1997). Social Network Effects on the Extent of Innovation Diffusion: A Computer Simulation. *Organization Science, Vol. 8, No. 3,* pp. 289-309.

Anderson, B.D.O., & Moore, J.B. (1979). Optimal Filtering. *1st Edition, Englewood Cliffs, N.J.: Prentice-Hall.*

Andrews, R.L. (1994). Forecasting Performance of Structural Time Series Models. *Proceedings of the 44th IEEE Conference on Decision Control,* pp. 332-337.

Andrieu, C., Doucet, A., & Tadic, V.B. (2005). On-Line Parameter Estimation in General State-Space Models. *Journal of Business & Economic Statistics, Vol. 12, No. 1,* pp. 129-133.

Bala, V., & Goyal, S. (1998). Learning from Neighbours. *The Review of Economic Studies, Vol. 65, No. 3,* pp. 595-621.

Bass, F.M. (1969). A New Product Growth Model for Consumer Durables. *Management Science, Vol. 15,* pp. 215-227.

Bass, F.M., Krishnan, T.V., & Jain, D.C. (1994). Why the Bass Model Fits Without Decision Variables. *Marketing Science, Vol. 13, No. 3,* pp. 203-223.

Bengtsson, T., & Cavanaugh, J.E. (2006). An Improved Akaike Information Criterion for State-Space Model Selection. *Computational Statistics & Data Analysis, Vol. 50,* pp. 2635-2654.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy, Vol. 100, No. 5,* pp. 992-1026.

Björk, Å(1996). Numerical Methods for Least Squares Problems. *1st Edition, Philadelphia: Society for Industrial and Applied Mathematics.*

Boswijk, H.P., & Franses, P.H. (2005). On the Econometrics of the Bass Diffusion Model. *Journal of Business & Economic Statistics, Vol. 23, No. 3,* pp. 255-268.

Box, G.E.P., & Jenkins, G.M. (1970). Time Series - Forecasting and Control. *1st Edition, San Francisco: Holden Day*.

Chandrasekaran, D., & Tellis, G. (2008). Global Takeoff of New Products: Culture, Wealth, or Vanishing Differences? *Marketing Science, Vol. 27, No. 5,* pp. 844-860.

Chandrasekaran, D., & Tellis, G. (2011). Getting a Grip on the Saddle: Chasms or Cycles? *Journal of Marketing, Vol. 75, No. 4,* pp. 21-34.

Chatterjee, K., & Xu, S.H. (2004). Technology Diffusion by Learning from Neighbors. *Advances in Applied Probability, Vol. 36, No. 2,* pp. 355-376.

Christensen, R., Johnson, W., Branscum, A., & Hanson, T.E. (2011). Bayesian Ideas and Data Analysis. *1st Edition, Boca Raton: CRC Press*.

Chu, C.-P., & Pan, J.-G. (2008). The Forecasting of the Mobile Internet in Taiwan by Diffusion Model. *Technological Forecasting & Social Change, Vol. 75, No. 7,* pp. 1054-1067.

Cojucaru, M.-G., Thille, H., Thommes, E., Nelson, D., & Greenhalgh, S. (2013). Social Influence and Dynamic Demand for New Products. *Environmental Modelling & Software, Vol. 50,* pp. 169-185.

Commandeur, J.J.F., & Koopman, S.J. (2007). An Introduction to State Space Time Series Analysis. *1st Edition, Oxford: Oxford University Press*.

Cowpertwait, P.S.P., & Metcalfe, A.V. (2009). Introductory Time Series with R. *1st Edition, New York: Springer*.

Davis, R.A., & Rodriguez-Yam, G. (2005). Estimation for State-Space Models Based on a Likelihood Approximation. *Statistica Sinica, Vol. 15,* pp. 381-406.

Delre, S.A., Jager, W., Bijmolt, T.H.A., & Janssen, M.A. (2010). Will It Spread or Not? The Effects of Social Influences and Network Topology on Innovation Diffusion. *Journal of Product Innovation Management, Vol. 27, No. 2,* pp. 267-282.

Dodds, W. (1973). An Application of the Bass Model in Longterm New Product Forecasting. *Journal of Marketing Research, Vol. 10, No. 3,* pp. 308-311.

Durbin, J., & Koopman, S.J. (2001). Time Series Analysis by State Space Methods. *Number 24 in Oxford Statistical Science Series, Oxford: Oxford University Press*.

Emmanouilides, C.J., & Davies, R.B. (2007). Modelling and Estimation of Social Network Effects in New Product Diffusion. *European Journal of Operational Research, Vol. 177,* pp. 1253-1274.

Frühwirth-Schnatter, S. (2006). Finite Mixture and Markov Switching Models. *1st Edition, New York: Springer*.

Fourt, L.A., & Woodlock, J.W. (1960). Early Prediction of Market Success for Grocery Products. *Journal of Marketing, Vol. 25,* pp. 31-38.

Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis, Vol. 1, No. 3,* pp. 515-534.

Goldenberg, J., Libai, B., & Muller, E. (2002). Riding the Saddle: How Cross-Market Communications Can Create a Major Slump in Sales. *Journal of Marketing, Vol. 66, No. 2,* pp. 1-16.

Golder, P.N., & Tellis, G.J. (1997). Will It Ever Fly? Modeling the Takeoff of Really New Consumer Durables. *Marketing Science, Vol. 16, No. 3,* pp. 256-270.

Golder, P.N., & Tellis, G.J. (2004). Growing, Growing, Gone: Cascades, Diffusion, and Turning Points in the Product Life Cycle. *Marketing Science, Vol. 23, No. 2,* pp. 207-218.

Golub, B., & Jackson, M.O. (2010). Nave Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics, Vol. 2, No. 1,* pp. 112-149.

Greene, W.H. (2012). Econometric Analysis. *7th Edition, Boston: Pearson*.

Gujarati, D.N. (2003). Basic Econometrics. *4th Edition, New York: McGraw-Hill*.

Guidolin, M., & Guseo, R. (forthcoming). Modelling Seasonality in Innovation Diffusion. *Technological Forecasting & Social Change*.

Gustafson, P., Hossain, S., & MacNab, Y.C. (2006). Conservative Prior Distributions for Variance Parameters in Hierarchical Models. *The Canadian Journal of Statistics, Vol. 34, No. 3,* pp. 377-390.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. *1st Edition, Cambridge: Cambridge University Press*.

Harvey, A.C. (1990). The Econometric Analysis of Time Series. *2nd Edition, London: Philip Allan*.

Harvey, A.C., & Peters, S. (1990). Estimation Procedures for Structural Time Series Models. *Journal of Forecasting, Vol. 9,* pp. 89-108.

Harvey, A.C., & Shephard, N. (1993). Structural Time Series Models. *Handbook of Statistics, Vol. 11: Econometrics (edited by G.S. Maddala, C.R. Rao, & H.D. Vinod), Amsterdam: North Holland,* pp. 261-302.

Harvey, A.C., & Todd, P.H.J. (1983). Forecasting Economic Time Series with Structural and Box-Jenkins Models. *Journal of Business and Economic Statistics, Vol. 1,* pp. 299-315.

Hyndman, R.J., & Koehler, A.B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting, Vol. 22, No. 4,* pp. 679-688.

Jackman, S. (2009). Bayesian Analysis for the Social Sciences. *1st Edition, Chichester: Wiley*.

Jackson, M.O. (2010). Social and Economic Networks. *1st Edition, Princeton NJ: Princeton University Press*.

Jackson, M.O., & Rogers, B.W. (2007). Meeting Strangers and Friends of Friends: How Random Are Social Networks? *American Economic Review, Vol. 97, No. 3,* pp. 890-915.

Jackson, M.O., & Yariv, L. (2007). Diffusion of Behavior and Equilibrium Properties in Network Games. *American Economic Review, Vol. 97, No. 2,* pp. 92-98.

Jazwinski, A. (1970). Stochastic Processes and Filtering Theory. *1st Edition, New York: Academic Press*.

Jeuland, A.P. (1995). The Bass Model as a Tool to Uncover Empirical Generalizations in Diffusion of Innovation. *Empirical Generalizations Conference, University of Chicago, Wharton School, 1995*.

Jiang, Z., Bass, F.M., & Bass, P.I. (2006). Virtual Bass Model and the Left-Hand Data-Truncation Bias in Diffusion of Innovation Studies. *International Journal of Research in Marketing, Vol. 23,* pp. 93-106.

Jukic, D. (2011). Total Least Squares Fitting Bass Diffusion Model. *Mathematical and Computer Modelling, Vol. 53,* pp. 1756-1770.

Jukic, D. (2013a). On Nonlinear Weighted Least Squares Estimation of Bass Diffusion Model. *Applied Mathematics and Computation, Vol. 219,* pp. 7891-7900.

Jukic, D. (2013b). On the l_s-norm Generalization of the NLS Method for the Bass Model. *European Journal of Pure and Applied Mathematic, Vol. 6, No. 4,* pp. 435-450.

Kapur, S. (1995). Technological Diffusion with Social Learning. *Journal of Industrial Economics, Vol. 43, No. 2,* pp. 173-195.

Kendall, M.G. (1973). Time Series. *1st Edition, London: Charles Griffin & Co.*

Kijek, A., & Kijek, T. (2010). Modelling of Innovation Diffusion. *Operations Research and Decisions, Vol. 3-4,* pp. 53-68.

Kitagawa, G., & Gersch, W. (1996). Smoothness Priors Analysis of Time Series. *1st Edition, New York: Springer.*

Koopman, S.J., Harvey, A.C., Doornik, J.A., & Shephard, N. (2009). STAMP 8.2: Structural Time Series Analyser, Modeler, and Predictor. *London: Timberlake Consultants.*

Kreindler, G.E., & Young, H.P. (2013). Rapid Innovation Diffusion in Social Networks. *Oxford Department of Economics Discussion Paper Series, No. 626.*

Kuandykov, L., & Sokolov, M. (2010). Impact of Social Neighborhood on Diffusion of Innovation S-Curve. *Decision Support Systems, Vol. 48,* pp. 531-535.

Link, W.A., & Eaton, M.J. (2012). On the Thinning of Chain in MCMC. *Methods in Ecology and Evolution, Vol. 3, No. 1,* pp. 112-115.

Lopez-Pintado, D., & Watts, D.J. (2008). Social Influence, Binary Decisions, and Collective Dynamics. *Rationality and Society, Vol. 20, No. 4,* pp. 399-443.

Mahajan, V., Muller, E., & Bass, F.M. (1990). New Product Diffusion Models In Marketing: A Review and Directions for Research. *Journal of Marketing, Vol. 54,* pp. 1-26.

Mahajan, V., Muller, E., & Bass, F.M. (1995). Diffusion of New Products: Generalizations and Managerial Uses. *Marketing Science, Vol. 14, No. 2,* pp. G79-G88.

Mahajan, V., Muller, E., & Wind, J. (2000). New Product Diffusion Models. *1st Edition, New York: Klower Academic Publishers.*

Mahajan, V., & Peterson, R.A. (1985). Models for Innovation Diffusion. *1st Edition, Beverly Hills: Sage Publications.*

Mahajan, V., & Sharma, S. (1986). A Simple Algebraic Estimation Procedure for Innovation Diffusion Models of New Product Acceptance. *Technological Forecasting & Social Change, Vol. 30,* pp. 331-345.

Mahajan, V., Srinivasan, V., & Mason, C.H. (1986). An Evaluation of Estimation Procedures for New Product Diffusion Models. In: *In: V. Mahajan, Y. Wind (Eds.), Innovation Diffusion Models of New Product Acceptance, Massachusetts, Ballinger Cambridge,* pp. 203-232.

Manuelli, R.E., & Seshadri, A. (2014). Frictionless Technology Diffusion: The Case of Tractors. *American Economic Review, Vol. 104, No. 4,* pp. 1368-1391.

Meade, N., & Islam, T. (2006). Modelling and Forecasting the Diffusion of Innovation A 25-Year Review. *International Journal of Forecasting, Vol. 22, No. 3,* pp. 519-545.

Mendelsohn, R. (2011). The STAMP Software for State Space Models. *Journal of Statistical Software, Vol. 41, No. 2,* pp. 1-18.

Michalakelis, C., Varoutas, D., & Sphicopoulos, T. (2008). Diffusion Models of Mobile Telephony in Greece. *Telecommunications Policy, Vol. 32, No. 3-4,* pp. 234-245.

Montanari, A., & Saberi, A. (2010). The Spread of Innovations in Social Networks. *Proceedings of the National Academy of Sciences, Vol. 107, No. 47,* pp. 20196-20201.

Moore, G.A. (1991). Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers.. *1st Edition, New York: Harper Collins*.

Morris, S. (2000). Contagion. *Review of Economic Studies, Vol. 67, No. 1,* pp. 57-78.

Muller, E., & Yogev, G. (2006). When Does the Majority Become a Majority? Empirical Analysis of the Time at Which Main Market Adopters Purchase the Bulk of Our Sales. *Technological Forecasting & Social Change, Vol. 73, No. 10,* pp. 1107-1120.

Munshi, K. (2004). Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution. *Journal of Development Economics, Vol. 76, No. 1,* pp. 185-213.

Naseri, M.B., & Elliott, G. (2013). The Diffusion of Online Shopping in Australia: Comparing the Bass, Logistic and Gompertz Growth Models. *Journal of Marketing Analytics, Vol. 1, No. 1,* pp. 49-60.

Norton, J. A., & Bass, F.M. (1987). A Diffusion Theory Model of Adoption and Substituion for Successive Generations of High-Technology Products. *Management Science, Vol. 33,* pp. 1069-1086.

OECD (2005). Oslo Manual - Guidelines for Collecting and Interpreting Innovation Data. *3rd Edition, OECD Publishing*.

Peres, R., Muller, E., & Mahajan, V. (2010). Innovation Diffusion and New Product Growth Models: A Critical Review and Research Directions. *International Journal of Research in Marketing, Vol. 27, No. 2,* pp. 91-106.

Petris, G., & Petrone, S. (2011). State Space Models in R. *Journal of Statistical Software, Vol. 41, No. 4,* pp. 1-25.

Petris, G., Petrone, S., & Campagnoli, P. (2009). Dynamic Linear Models in R. *1st Edition, London: Springer*.

R Development Core Team (2012). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.*.

Rodgers, E.M. (2003). Diffusion of Innovations. *5th Edition, New York: Free Press*.

Sage, A.P., & Melsa, J.L. (1971). System Identification. *1st Edition, New York: Academic Press*.

Satoh, D. (2001). A Discrete Bass Model and Its Parameter Estimation. *Journal of the Operations Research Society of Japan, Vol. 44,* pp. 1-18.

Schmittlein, D.C., & Mahajan, V. (1982). Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance. *Marketing Science, Vol. 1, No. 1,* pp. 57-78.

Scitovski, R., & Meler, M. (2002). Solving Parameter Estimation Problem in New Product Diffusion Models. *Applied Mathematics and Computation, Vol. 127, No. 1,* pp. 45-63.

Shankar, V., & Carpenter, G.S. (2012). Handbook of Marketing Strategy. *1st Edition, Cheltenham: Edward Elgar*.

Sharif, M.N., & Ramanathan, K. (1981). Binomial Innovation Diffusion Models With Dynamic Potential Adopter Population. *Technological Forecasting & Social Changes, Vol. 20,* pp. 63-87.

Sneek, J.M. (1984). Modelling Procedures for Economic Time Series. *1st Edition, Amsterdam: Free University Press.*

Srinivasan, V., & Mason, C.H. (1986). Nonlinear Least Squares Estimation of New Product Diffusion Models. *Marketing Science, Vol. 5,* pp. 169-178.

Sterman, J. (2000). Business Dynamics: Systems Thinking and Modelling for a Complex World. *1st Edition, Boston: Irwin McGraw-Hill.*

Stremersch, S., & Tellis, G.J. (2004). Understanding and Managing International Growth of New Products. *International Journal of Research in Marketing, Vol. 21, No. 4,* pp. 421-438.

Strickland, C.M., Martin, G.M., & Forbes, C.S. (2008). Parameterisation and Efficient MCMC Estimation of Non-Gaussian State Space Models. *Computational Statistics & Data Analysis, Vol. 52,* pp. 2911-2930.

Studenmund, A.H. (2006). Using Econometrics: A Practical Guide. *5th Edition, Boston: Pearson Addison Wesley.*

Sultan, F., Farley, J.U., & Lehmann, D.R. (1996). Reections on A Meta-Analysis of Applications of Diffusion Models. *Journal of Marketing Research, Vol. 33, No. 2,* pp. 247-249.

Tanizaki, H. (2001). Estimation of Unknown Parameters in Nonlinear and Non-Gaussian State-Space Models. *Journal of Statistical Planning and Inference, Vol. 96,* pp. 301-323.

Tanizaki, H., & Mariano, R.S. (1998). Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Simulations. *Journal of Econometrics, Vol. 83,* pp. 263-290.

Teyssiere, G. (2005). Structural Time Series Modelling with STAMP 6.02. *Journal of Applied Econometrics, Vol. 20,* pp. 571-577.

Valente, T. (1995). Network Models of the Diffusion of Innovations. *1st Edition, Cresskill NJ: Hampton Press.*

Van den Bulte, C., & Joshi, Y.V. (2007). New Product Diffusion with Influentials and Imitators. *Marketing Science, Vol. 26, No. 3,* pp. 400-421.

Van den Bulte, C., & Lilien, G. L. (1997). Bias and Systematic Change in the Parameter Estimates of Macro-Level Diffusion Models. *Marketing Science, Vol. 16, No. 4,* pp. 338-353.

Van den Bulte, C., & Stremersch, S. (2004). Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test. *Marketing Science, Vol. 23, No. 4,* pp. 530-544.

Wei, W.W.S. (1990). Time Series Analysis: Univariate and Multivariate Methods. *1st Edition, Wokingham: Addison-Wesley*.

West, M., & Harrison, J. (1997). Bayesian Forecasting and Dynamic Models. *1st Edition, New York: Springer*.

Wright, M., Upritchard, C., & Lewis, T. (1997). A Validation of the Bass New Product Diffusion Model in New Zealand. *Marketing Bulletin, Vol. 8,* pp. 15-29.

Wu, F.-S. & Chu, W.L. (2010). Diffusion Models of Mobile Telephony. *Journal of Business Research, Vol. 63, No. 5,* pp. 497-501.

Young, H.P. (2003). The Diffusion of Innovations in Social Networks. *The Economy as an Evolving Complex System, Vol. 3 (edited by Blume, L.E., & Durlauf, S.N.), Oxford: Oxford University Press,* pp. 267-282.

Young, H.P. (2005). The Spread of Innovations through Social Learning. *Brookings Institution Center for Social and Economic Dynamics, Working Paper No. 43*.

Young, H.P. (2009). Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning. *American Economic Review, Vol. 99,* pp. 1899-1924.

Young, P. (1993). Technological Growth Curves: A Competition of Forecasting Models. *Technological Forecasting & Social Change, Vol. 44, No. 4,* pp. 375-389.

Young, P., & Ord, J.K. (1989). Model Selection and Estimation for Technological Growth Curves. *International Journal of Forecasting, Vol. 5, No. 4,* pp. 501-513.

Appendix A - Analytical Solution of the Logistic Differential Equation

Starting from the logistic differential equation

$$\frac{dN(t)}{dt} = r \times N(t) \times (1 - \frac{N(t)}{m})$$

we perform a separation of variables:

$$\int \frac{dN(t)}{N(t) \times (1 - N(t)/m)} = \int r\,dt$$

To evaluate the left-hand side, we can write:

$$\frac{1}{N(t) \times (1 - N(t)/m)} = \frac{m}{N(t) \times (m - N(t))} = \frac{1}{N(t)} + \frac{1}{m - N(t)}$$

Hence,

$$\int \frac{dN(t)}{N(t)} + \int \frac{dN(t)}{m - N(t)} = \int r\,dt$$

$$ln|N(t)| - ln|m - N(t)| = rt + C$$

$$ln|\frac{m - N(t)}{N(t)}| = -rt - C$$

$$|\frac{m - N(t)}{N(t)}| = e^{-rt-C}$$

$$\frac{m - N(t)}{N(t)} = A \times e^{-rt} \qquad (A = \pm e^{-C})$$

From here we can solve for $N(t)$:

$$N(t) = \frac{m}{1 + A \times e^{-rt}}$$

where $A = \frac{m-N(0)}{N(0)}$. This is the solution that was used for the estimation of the logistic growth model in the main paper.